Sara Filipa Rato Bispo

Analysis of glycosylation heterogeneity in antibody production by *Pichia pastoris*

Lisboa

2009

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Departamento de Química

Analysis of glycosylation heterogeneity in antibody production by *Pichia pastoris*

Sara Filipa Rato Bispo

Dissertação apresentada na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa para obtenção do grau de Mestre em Biotecnologia.

Tese orientada por: Professor Doutor Rui Oliveira

Lisboa

2009

ACKNOWLEDGMENTS

Em primeiro lugar queria agradecer à minha família em especial aos meus pais pelo carinho e toda a compreensão, aos meus avós e ao meu querido irmão por todo o apoio.

Queria agradecer ao Prof. Rui Oliveira pela orientação, à Ana Ferreira pela ajuda e apoio, ao João Dias, e a todos da Unidade Piloto do Instituto de Biologia Experimental e Tecnológica.

A todos os meus amigos e em especial à Ana Catarina por viver comigo todos os momentos, bons e maus.

Ao meu namorado e amigo, Marco, por partilhar a minha vida sempre com uma palavra de motivação e encorajamento.

A todos um muito Obrigada!

ABSTRACT

The aims of this M.Sc. thesis are the production of an antibody fragment in *Pichia pastoris* and to assess O-glycosylation heterogeneity theoretically and experimentally. All possible glycoforms (glycans that are attached to proteins) attached to this antibody were calculated and a simple mathematical model was developed in MATLAB to predict glycoforms heterogeneity.

The production of Anti-(ED-B) scFv by *Pichia pastoris* was made in a 50 L fed-batch fermenter. The maximum product concentration obtained at approximately 100 hours of operation was 8,4mg/L.

The Anti-(ED-B) scFv has no sequence of N-linked glycosylation and for the sequence of O-linked glycosylation there are 56 possible spots. Two mathematical models were developed in MATLAB: a deterministic and a stochastic model. Both models predict a given glycoforms distribution under the premise that the Endoplasmatic Reticulum enzymes and *Golgi* enzymes (O-Mannosyltranferase, α-1,2-Mannosyltransferase and β-1,2-Mannosyltransferase) are active. The models predict when the ratio of protein concentration per initial mannose concentration is low the prevalent glycoforms are protein with two mannoses. While when the ratio of protein concentration per initial mannose concentration is very low, the prevalent glycoforms are protein with five mannoses. For high number of mannose molecules at the beginning of the glycosylation process there is a convergence of results predicted by the deterministic and stochastic models. However, both models diverge when the number of mannose molecules is very low. In this situation, the stochastic predictions are more consistent with the true nature of the system than the deterministic ones.

Despite of a theoretical O-linked glycoforms distribution, experimental test by Glycoprotein Detection Kit showed that the anti-(ED-B) scFv produced by *Pichia pastoris* is not glycosylated at the tested culture conditions.

The posttranslational modifications that lead to the formation of monomers or dimers are not related to glycosylation but probably to the folding process.

In future studies it would be interesting to study if the occurrence of O-glycosylation and its heterogeneity can be controlled by reactor operation parameters such as Temperature, pH and methanol feeding rate. It is known that such operational parameters have an important impact on the conformation of the scFv.

RESUMO

Os objectivos desta tese de mestrado são produzir um fragmento de anticorpo na levedura *Pichia pastoris* e avaliar teoricamente e experimentalmente a heterogeneidade da O-glicosilação. Todas as possíveis glicoformas (glicanos que estão ligados às proteínas) que estão ligadas ao anticorpo foram calculadas e desenvolveu-se em MATLAB modelos matemáticos simples para prever a heterogeneidade das glicoformas.

A produção do scFv Anti-(ED-B) na *Pichia pastoris* foi efectuada num fermentador *fed-batch* de 50 L. A concentração máxima de produto obtida a aproximadamente 100h foi de 8,4mg/L.

O Anti-(ED-B) scFv não possui sequência para N-glicosilação e para a O-glicosilação existem 56 sítios possíveis. Foram desenvolvidos dois modelos matemáticos em MATLAB: o determinístico e o estocástico. Ambos os modelos prevêem uma distribuição de glicoformas sob a premissa que as enzimas do Retículo Endoplasmático e do *Golgi* (O-Mannosyltranferase, α-1,2-Mannosyltransferase and β-1,2-Mannosyltransferase) estão activas. Os modelos estimam que quando o rácio da concentração de proteína por concentração de manose inicial é baixo, a glicoforma predominante é proteína com uma manose. Enquanto que quando o rácio da concentração de proteína por concentração de manose é ainda menor, a glicoforma predominante é proteína com cinco manoses. Para números elevados de moléculas manose no inicio do processo de glicosilação existe uma convergência dos resultados previstos segundo os modelos determinístico e estocástico. Contudo, ambos os modelos divergem quando o número de moléculas manose é muito baixo. Neste contexto, o modelo estocástico é mais coerente com a verdadeira natureza do sistema que o modelo determinístico.

Apesar da distribuição teórica das glicoformas *O-linked*, o kit laboratorial de detecção de glicoproteínas indica que o anti-(ED-B) scFv produzido pela *Pichia pastoris* não é glicosilado, nas condições da cultura efectuada.

As modificações pós-traducionais que levam à formação dos monómeros ou dímeros não estão relacionadas com a glicosilação mas estão provavelmente ligadas com o processo de *folding.*

Em estudos futuros seria interessante estudar se a ocorrência de O-glicosilação e a sua heterogeneidade poderia ser controlada pelos parâmetros operacionais do reactor, tais como Temperatura, pH e taxa de alimentação do

metanol. Sabe-se que os parâmetros operacionais têm um impacto importante na conformação do scFv.

ABBREVIATION

1,2-ManT	α-1,2-Mannosyltransferase
1,6-ManT	α-1,6-Mannosyltransferase
A	Stochiometric matrix
ADP	Adenosine-5'-diphosphate
Age	Time in hours since the inoculum was transferred into the pilot fermenter
Asn	Asparagine
ATP	Adenosine-5'-triphosphate
BCA	Bicinchoninic acid
BSA	Bovine serum albumin
C_i	Concentration of the several compounds
C_P	Protein concentration
C_M	Mannose concentration
C_{PM}	Concentration of protein with one mannose
C_{PM_2}	Concentration of protein with two mannoses
C_{PM_3}	Concentration of protein with three mannoses
C_{PM_4}	Concentration of protein with four mannoses
C_{PM_5}	Concentration of protein with five mannoses
CHCA	α-ciano-4-hydroxycinaminic acid
Dol-P	Dolichol phosphate
E.R.	Endoplasmic reticulum
E.R. α-Man	E.R. α-mannosidase
ED-B	Extra domain B
ELISA	Enzyme-Linked Immunosorbent Assay
EndoF	Endoglycosidase F
EndoH	Endoglycosidase H
ERAD	ER-associated protein degradation
Gal	Galactose
GDP	Guanosine-5'-diphosphate
Glc	Glucose
Glc I	α-glucosidase I
Glc II	α-glucosidase II
GlcNac	N-acetylglucosamine
GTP	Guanosine-5'-triphosphate

HRP	Horseradish peroxidase
M_0	Initial concentration of mannoses
Man	Mannose
MALDI	Matrix-assisted laser desorption/ionization
MS	Mass spectrometry
MUT	Methanol UTilization
Neu5Ac	N-Acetylneuraminic acid
Neu5Ac	N-Acetylneuraminic acid
OD	Optical density
OPD	o-Phenylenediamine dihydrochloride
OST	N-Oligosaccharyl transferase
P	Protein non-glycosylated
PAS	Periodic Acid-Schiff
PBS	Phosphate buffered saline
PM_1	Protein with one mannose
PM_2	Protein with two mannoses
PM_3	Protein with three mannoses
PM_4	Protein with four mannoses
PM_5	Protein with five mannoses
PNGase	Peptide N-Glycosidase F
Q	Feeding rate (g/h)
$Q_{0,\ Gly}$	Initial feeding rate (g/h)
RNA	Ribonucleic acid
r_i	Kinetic constant
scFv	Single-chain antibody of variable fragment
SDS-PAGE	SodiumDdodecyl Sulfate PolyAcrylamide Gel Electrophoresis
Ser, S	Serine
SRP	Signal recognition particle
t_r	Residence time
Thr, T	Threonine
TOF	Time-of-flight
UDP	Uridine diphosphate
μ_1 / μ_2	Desired specific growth rate
V_H	Variable heavy domain
V_L	Variable light domain
WCW	Wet cell weight

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

The aim of this study is to produce an antibody fragment in *Pichia pastoris*, applying the knowledge of metabolic modeling (Jahic, *et al*., 2002) (this isn't the principal aim so the data isn't provided); to find out the possible glycoforms (glycans that are attached to proteins) attached to this antibody and to develop some technique to detect and quantify the glycoforms (Morelle, *et al*., 2005), since the glycosylation can affect the properties of the proteins (Leibiger, *et al*., 1999); and to begin some programs to visualize the distribution of glycofoms in the proteins (Shelikoff, *et al*., 1996; Hossler, *et al*., 2006). It is important to find techniques that are able to manipulate the production of proteins with correct glycosylation, for that to happen it is necessary to know correctly how the glycoforms are produced and have this information in the metabolic network (Gonzalez, *et al*., 2001; Umaña, *et al*., 1997).

The introduction begins to explain the protein itself, a brief explanation of the production process and ends with description of glycosylation in *Pichia pastoris*, since this information is hard to find and it is very important to know, case in future, this information is going to be use in metabolic modeling.

1.1. Protein of interest

In the last decades, many monoclonal antibodies were studied with the purpose of therapeutic applications, in particular cancer treatments.

The protein of interest in this study is Anti-ED-B Fibronectin antibody fragment.

The fibronectin is a universal cell-adhesion molecule which has an alternative splicing. This splice variants is ED-B fibronectin. This molecule is undetectable in normal adult tissue, although it exhibits great expression in fetal and tumor tissues. This protein is also associated with tissue remodeling during a wound healing and with neovasculature solid tumor (like invasive ductal carcinoma, aggressive brain tumor and ocular angiogenesis) (Berndorff, *et al*., 2006; Philogen®). During angiogenesis the ED-B fibronectin is accessible from the blood stream – so these properties make it a perfect marker for angiogenesis (Berndorff, *et al*., 2005).

As result, it is very important to develop a target to ED-B fibronectin. Some of the studies begin to use full immunoglobulins but the long circulation time causes high levels of toxicity to healthy tissues. Therefore, it is important to find molecules that can

identify and accumulate into the tumor and be eliminated from the blood stream, not causing toxicity to healthy tissues (Berndorff, *et al*., 2005). That molecule can be a single-chain antibody of variable fragment (scFv) and in our case of study is the anti-ED-B fibronectin antibody fragment – anti-ED-B scFV.

The purpose of develop this antigen is to fuse him to a cytokine (a signaling molecule). This aggregate will target the ED-B fibronectin which are at the tumor site – therefore it localizes efficiently the tumor. If the cytokine has some properties to treat or to keep apart the tumor from the rest of the healthy cells, the aggregate accomplish two different and very important tasks (Philogen®).

1.2. Expression system

The production of recombinant proteins has been studied in prokaryotic and eukaryotic expression systems.

The production of eukaryotic proteins in bacteria has problems due to instability or lack of biological activity; and, in some cases, bacterial compounds can cause several toxicity reactions in humans and in animals, despite the careful purification methods.

If the aim is for medical use, the recombinant proteins must be identical to the natural proteins in all its proprieties. For this to occur, it is necessary that the expression systems are able to accomplish all of the post-translational modifications; some examples are the correct disulfide bonds and the glycosylation process.

The *Pichia pastoris* expression system is a very popular system because it has a number of advantages in comparison with other yeast systems: the expression vector is highly efficient and tightly regulated; it produces large amounts of protein; and, normally, it secretes very few proteins (including proteolytic enzymes with the ability to degrade product), which it is an advantage for the purification process (Glick, *et al*., 2003).

Pichia pastoris is a methylotrophic yeast; it is regulated by the promoter of the methanol inducible gene that encodes for alcohol oxidase – the first enzyme of the methanol pathway.

A typical MUT (Methanol UTilization) *P. pastoris* expression vector is generally shown in Figure 1.1. Most of these vectors are designed to be integrated in *P. pastoris*

chromosome to avoid strain instability problems and to promote a long term growth process (Invitrogen®).

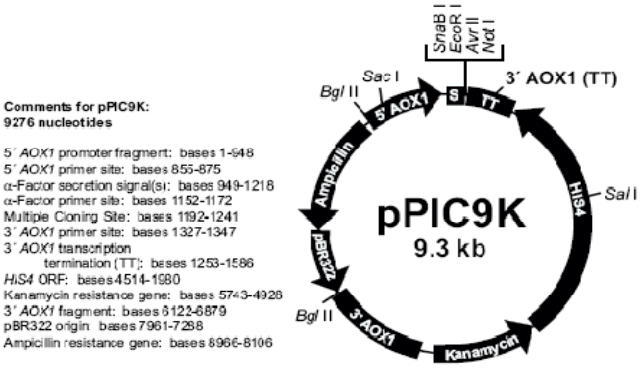

Figure 1.1 – Representation of the pPIC9K vector (Invitrogen®). The gene of the target foreign protein is placed in the multiple cloning site.

The regulation is provided by AOX gene – at transcriptional level – when methanol is present in the medium, it triggers the transcription of the foreign gene. In the absence of methanol, the expression of that gene is completely turned off. This promoter enables large-scale production of high amounts of recombinant protein, since it's a strong promoter that responds rapidly to the presence/absence of methanol in medium (Glick, *et al*., 2003).

1.3. Fermentation

Optimal cultivation strategies for the production of heterologous protein in *Pichia pastoris* have been developed by many researchers (e.g. Cunha, *et al*., 2004, Damasceno, *et al*., 2004, Hellwig, *et al*., 2001, Cos, *et al*., 2006, d'Anjou, *et al*., 2000). Most of these studies aim at maximizing product titer at the end of the cultivation. There are no works in the literature that study the effect of cultivation conditions on glycosylation patterns in *P. pastoris*.

In a typical high cell density MUT *P. pastoris* strain two different carbon sources are utilized, namely glycerol and methanol.

In the first and second phase glycerol is fed into the reactor with the goal of biomass production. In a third phase, i.e. the post-induction phase, methanol is fed to the reactor, to induce the production of the scFv and also to serve as carbon source for cell growth and product production (Cos, *et al*., 2006).

As stated previously (1.2. Expression system), the expression of scFv is regulated by the alcohol oxidase promoter. Thus the presence of methanol triggers the production of the protein. However, the feeding of methanol is not straightforward since the methanol is both an inhibitor and an inducer substrate and at the same time could be inhibitory for the cellular growth (methanol concentration 65g/L). The decision of the methanol feeding strategy in the post-induction phase is thus a critical point in MUT *P. pastoris* cultures (Cunha, *et al*., 2004; d'Anjou, *et al*., 2000).

In the transition between the glycerol phase to the methanol phase there is a short moment in time with both methanol and glycerol present in the medium. A small injection of methanol (~1 g/l) is fed before the end of the glycerol exponential feeding. This procedure has the purpose to stimulate the synthesis of alcohol oxidase and consequently accelerate the adaption to methanol feeding in the second stage.

For maximizing the production of both cellular growth and protein production other parameters like temperature, pH, oxygen levels and agitation need to be considered. The conditions of all of these parameters and the feeding of the carbon source can shape the synthesis, processing, secretion of the protein and the cellular growth; and change some specific characteristics of the protein produced (Cos, *et al*., 2006; Cunha, *et al*., 2004). One still open issue is to understand the controllability of glycosylation patterns by manipulating such operational parameters.

1.4. Glycosylation

Pichia pastoris combines the simple cellular maintenance like the lower organisms, such as bacteria, with the posttranslational processing machinery of eukaryotes (Cos, *et al*., 2006).

Glycosylation is one modification that can occur in protein biosynthesis, after or during the translation. Others modifications that occur are:

- Formation of disulfide bonds;
- Proper folding;
- Specific proteolytic cleavages;
- Assembly into multimeric proteins (Lodish, *et al*., 1999).

Before developing this topic, it is important to mention that glycosylation in *Pichia pastoris* is very similar to glycosylation in other yeast species, like

Saccharomyces cerevisae, obviously the yeast in question is not engineered to perform a particular type of glycosylation. So, along the text it'll be mentioned if the mechanisms are general in yeast; or if it they are particular of the one or the other species.

The polypeptides chain that have to walkthrough the secretory pathway begin their path in the Endoplasmic reticulum (E.R.), pass in *Golgi* apparatus, form a vacuole or a lysosome and the final step is crossing the plasma membrane to the extracellular medium (Corsi, *et al*., 1996).

When the polypeptides are being translated they have an E.R. signal sequence normally localized in N-terminus. This sequence is typically formed by one or more positively charged amino acids followed by continuous sequence of 6-12 hydrophobic residues (Lodish, *et al*., 1999).

The translocation, crossing through E.R. membrane, has two possible pathways in yeast, including *Pichia pastoris*. The first one is Co-translational targeting: at the same time the polypeptide is being translated, it is also being transferred for E.R. lumen; the second one is Post-translational: first the polypeptide is translated and only after that it is transferred to E.R. lumen. It appears that the latter pathway is more frequently used (by a larger fraction of intracellular proteins) because the process of translation doesn't always happen close to the E.R. (Corsi, *et al*., 1996; Rapoport, 2007).

Co-translational targeting

The signal sequence of the polypeptide chain is identified by the signal recognition particle (SRP) which is a ribonucleoprotein composed of six polypeptides and a 300-nucleotide RNA. The SRP binds to the signal sequence when it emerges from the ribosome and it is important to mention that this bioreaction has a consumption of GTP (Guanosine-5'-triphosphate – one of the molecules for energy in the biological systems). This ribosome-nascent-chain-SRP complex is recognized by SRP receptor – a protein of E.R. membrane. This SRP receptor takes the complex and places it in a translocon (a transmembrane protein). When the ribosome is attached to the translocon, the elongating of polypeptide chain starts inside the tunnel of the ribosome into channel of the translocon. Then, the SRP is released and recycled back into the cytoplasm (Corsi, *et al*., 1996). All these steps are illustrated in the schematic representation of Figure 1.2.

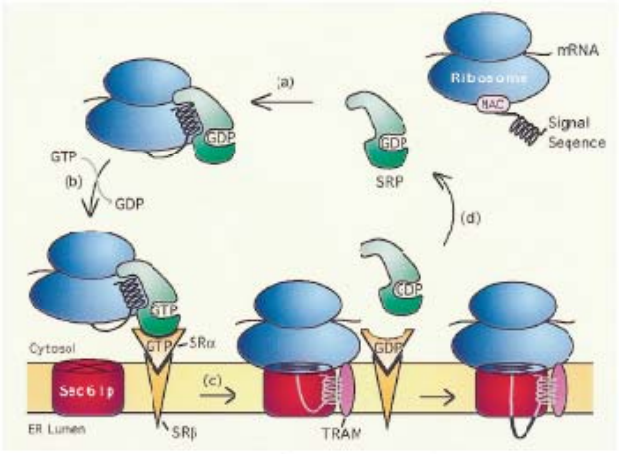

Figure 1.2 - Model for Co-translational targeting (Corsi, *et al*., 1996).

Post translational targeting

After translation the polypeptide chain is protected by cytosolic chaperones. The chaperones are proteins which help the polypeptide chain to remain unfolded and prevent the attack of proteases.

The Sec62/Sec63 complex recognizes the polypeptide chain and helps to place the signal sequence into Sec61 complex (translocon) (Figure 1.3). Then the translocation begins, the polypeptide chain enters into the channel of the translocon. At the same time, all of cytosolic chaperones are released from the polypeptide chain. When the chain starts to enter the E.R., the BiP (luminal chaperone) attaches the polypeptide chain to prevent their return to the cytosol. For the BiP to bind to the polypeptide chain, it has to be attached with an ATP and this joint must interact with J-domain of Sec63. During the binding, an ATP hydrolysis takes place. With the entrance of the polypeptide chain more bindings of other molecules of BiP occur. This process is repeated until the chain is completely inside of the E.R.. In the end, the ADP is converted to ATP and BiP is released from the polypeptide chain (Rapoport, 2007).

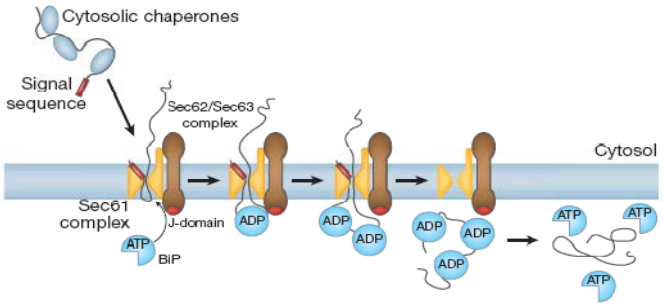

Figure 1.3 – Model for Post translational targeting (Rapoport, 2007).

1.4.1. Processing of N-linked Oligosaccharide

Assembly and Transfer of N-linked Oligosaccharide precursor

Before the N-linked Oligosaccharide is transferred to the protein that enters the E.R., this molecule needs to be synthesized.

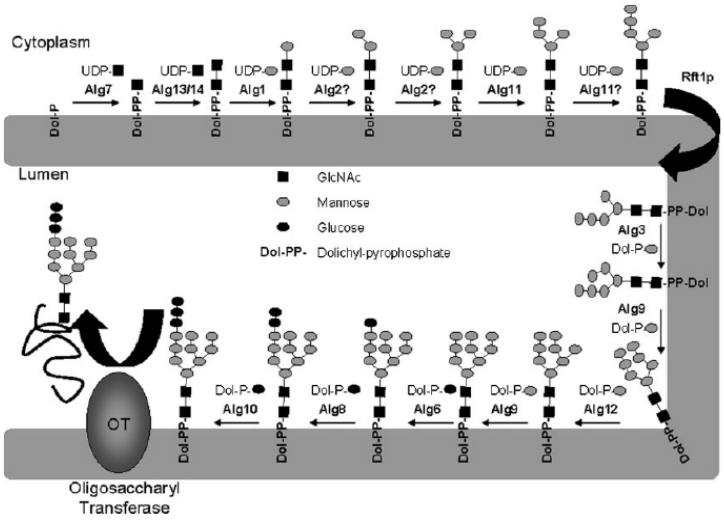

Figure 1.4 – Model of dolichol pathway of N-linked protein glycosylation in *Saccharomyces cerevisiae* (Weerapana, *et al*., 2006)

In higher eukaryotic systems the dolichol pathway is highly conserved (Figure 1.4): the glycan is transferred into asparagine site chain ($Glc_3Man_9GlcNac_2$β1,N-Asn) within Asn-X-Ser/Thr sequence (where X is any aminoacid except proline) (Kukuruzinska, *et al*., 1998; Helenius, *et al*., 2004; Weerapana, *et al*., 2006).

The first seven steps of the precursor assembly occur in the membrane of the E.R. at the cytosolic site. One of the limitations of this process is the lack of availability

of UDP-Man and UDP-GlcNac in the cytosol. After that, the heptasaccharide is flipped into the E.R. lumen with the membrane-spanning flippase (Rft1p). This enzyme is bi-directional and ATP-independent.

The remaining steps are performed with mannosyltransferase which transfers mannose from Dol-P-Man and glucosyltransferase which transfers glucose from Dol-P-Glc (Weerapana, *et al*., 2006). The limitations in these steps are the formation of Dol-P-Man and Dol-P-Glc in the E.R. at the cytosolic site and their flip to the E.R. lumen (Burda, *et al*., 1999).

It is important to refer that the dolichol phosphate (Dol-P) is recycled, in all steps of this procedure.

Once the precursor is formed, it will be transferred to the protein by the N-Oligosaccharyl transferase complex (OST). This complex is a multimeric, membrane associated enzyme, with an active site disposed into ER lumen.

In E.R., after the transfer of N-linked oligosaccharides, these glycans are trimmed – Figure 1.5 (Bretthauer, *et al*., 1999; Kukuruzinska, *et al*., 1987; Vervecken, *et al*., 2004). The three glucose residues are removed by α-glucosidase I for the first glucose residue, and α-glucosidase II for the next two glucose residues. The α-1,2-linked mannose is trimmed by E.R. α-mannosidase. So at this point, the protein has the N-linked $Man_8GlcNAc_2$ and it is prepared to follow the path for the *Golgi* apparatus to complete the transformation.

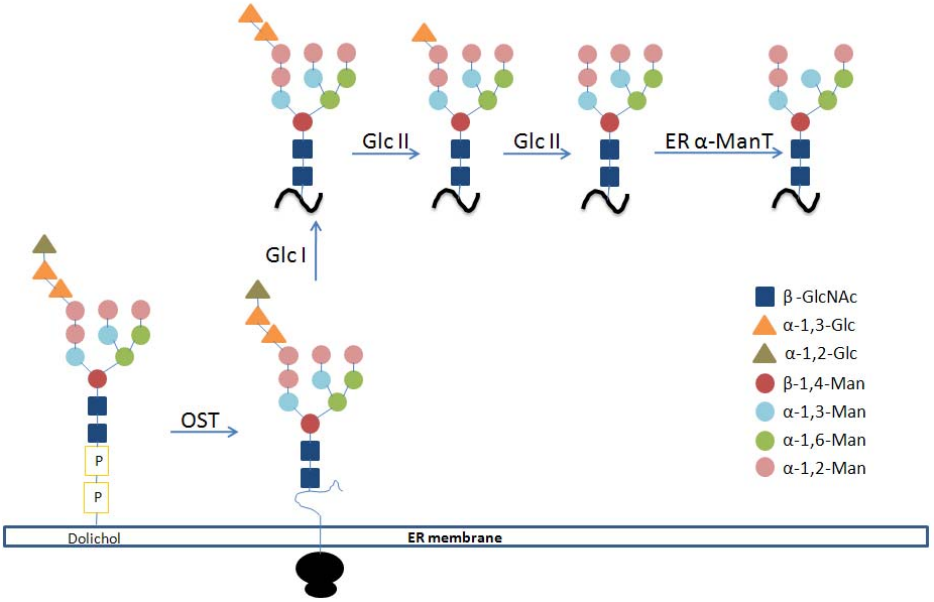

Figure 1.5 – Pathway of N-glycosylation in E.R. of *Saccharomyces cerevisae*. OST – N-Oligosaccharyl transferase; Glc I - α-glucosidase I; Glc II α-glucosidase II; ER α-Man - ER α-mannosidase. Adapted from (Herscovics, 1999).

Passing from ER to Golgi apparatus

If the protein has the correct fold, it is transported to the *Golgi* apparatus; otherwise it goes to the ER-associated protein degradation (ERAD) pathway. The folded-glycoprotein is transported to *Golgi* in a vesicle (Lee, *et al*., 2004).

It is important to refer that the *Golgi* apparatus in *Pichia pastoris* has three to five separate *Golgi* stacks, each of them adjacent to ER. In this respect, *P. pastoris* is very similar to mammalian *Golgi*, unlike *Saccharomyces cerevisiae* which has a disperse *Golgi* (Mogelsvang, *et al*., 2003, Rossanese, *et al*., 1999).

Transformation in Golgi apparatus

It is in this step where *Pichia pastoris* differs from mammal cells. In mammal cells the $Man_8GlcNAc_2$ glycans are trimmed to $Man_5GlcNAc_2$ and then more monosaccharides are added (like Gal, GlcNAc, and Neu5Ac).

In yeast, the $Man_8GlcNAc_2$ glycans aren't trimmed but usually elongated further, sometimes forming hypermannosyl glycans (Vervecken, *et al*., 2004). This

hypermannosyl glycans could be one of the restrictions for complete molecular characterization of the glycoproteins.

In *Pichia pastoris Golgi*, the only enzymes for glycosylation processing that are active are: α-1,2-Mannosyltransferase (1,2-ManT) and α-1,6-Mannosyltransferase (1,6-ManT), which transfer α-1,2-Mannose and α-1,6-Mannose, respectively (Figure 1.6). No α-1,3-Mannose linkages are observed in *Pichia pastoris*, this meaning that *P. pastoris* doesn't have or doesn't utilize α-1,3-Mannosyltransferase in the glycosylation (Bretthauer, *et al*., 1999).

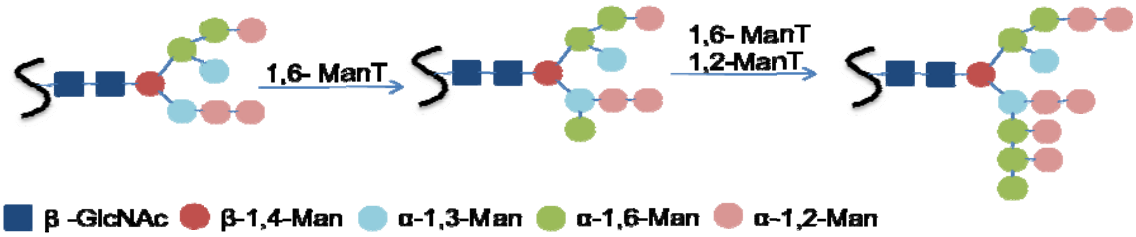

Figure 1.6 - Pathway of N-glycosylation in Golgi of Pichia pastoris. 1,2-ManT - α-1,2-Mannosyltransferase; 1,6-ManT - α-1,6-Mannosyltransferase. Adapted from (Hamilton, *et al*., 2007)

1.4.2. Processing of O-linked Oligosaccharide

The biosynthesis of O-glycosylation in comparison with N-glycosylation is relatively simple. The O-linked oligosaccharide chain can be attached to serine (Ser, S) or threonine (Thr, T) of the polypeptide chain.

As in N-glycosylation, the synthesis of O-glycans begins in the E.R. and after that in the *Golgi.*

In E.R., the first mannose residue is transferred from Dol-P-Man (with dolichol-phosphate-mannose protein mannosyltyransferase). This process has a limitation step: formation of Dol-P-Man in E.R. at cytosolic site – the lack of availability of mannose in cytosol (this limitation is very similar to the limitation in N-glycosylation) (Herscovics, *et al*., 1993).

The first and second mannose residues are transferred in the E.R. while the subsequent mannoses are transferred in the *Golgi* (Figure 1.7). The second and subsequent mannose residues are transferred directly from GDP-Man (Herscovics, *et al*., 1993).

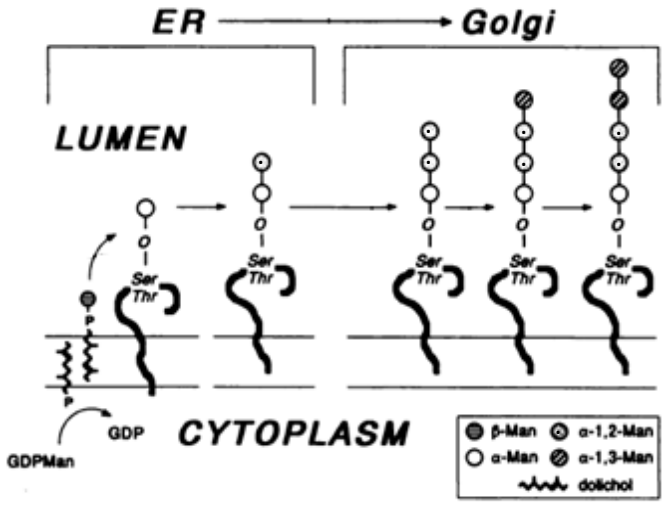

Figure 1.7 – Pathway of O-glycosylation in E.R. and *Golgi* of *Saccharomyces cerevisiae* (Herscovics, *et al*., 1993)

In yeast, and in particularly in *Pichia pastoris*, the structures of O-linked Oligosaccharides can have a maximum of five mannoses, represented in Figure 1.8 (Goto, 2007).The enzymes involved are O-mannosyltranferase for the first mannose, α-1,2-Mannosyltransferase for the second and third mannoses and β-1,2-Mannosyltransferase for the fourth and fifth mannoses.

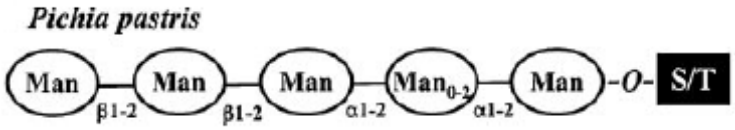

Figure 1.8 – Structure of O-glycans from *Pichia pastoris* (Goto, 2007)

2. Materials and Methods

2.1. Culture Mediums

For the pre-inoculum and inoculum: H_3PO_4 (85%) 26,7ml; $CaSO_4$ 0,93 g/L; K_2SO_4 18,2 g/L; $MgSO_4.7H_2O$ 14,9 g/L; KOH 4,13 g/L; $C_3H_5(OH)_3$ 40 g/L plus the solution of trace metals.

The solution of Trace metals: $CuSO_4.5H_2O$ 6 g/L; NaI 0,08 g/L; $MnSO_4.H_2O$ 3 g/L; $Na_2MoO_4.2H_2O$ 0,2 g/L; H_3BO_3 0,2 g/L; $CoCl_2$ 0,5 g/L; $ZnSO_4$ 20 g/L; $FeSO_4.7H_2O$ 65 g/L; biotin 0,2 g/L; H_2SO_4 5 ml.

The medium for the fermentation is the same than the inoculum except that it doesn't contain the solution of Trace metal and contains an antifoam, SAG 471, 0,2 g/L.

2.2. Fermentation

A 125ml T-flask with 40ml of BSM medium was inoculated with one cryovial from *Pichia pastoris* cell bank and incubated for 72 to 90 hours at 30ºC with agitation of 150rpm. A 2L flask with 750ml of BSM medium was inoculated with 10ml of the pre-inoculum – this shake flask culture had the duration of 72 hours at 30ºC with agitation of 150rpm. The resulting inoculum was finally transferred to the pilot fermenter with 15L of BSM medium – starting volume. The fermentation was conducted up to a maximum of 30L working volume in a pilot 50L fermenter (50L Laboratory Pilot Fermenter Type LP351, Bioengineering AG, Wald, Switzerland).

As described in the introduction, the fermentation had three phases with the control parameters changing from phase to phase.

The airflow rate was kept constant during all three phases, at 30 l/min.

The temperature during phase I and phase II are constant at 30ºC; in phase III until 50h follow the Equation 2.1, after 50h temperature is constant at 30ºC.

$$Temp = 0{,}1973\ x\ Age + 20{,}0319 \qquad \text{Equation 2.1}$$

with Age – time in hours since the inoculum was transferred into the pilot fermenter .

For pH the phase I and II are controlled to 5,0 with ammonium hydroxide solution; the phase III until 50h the Equation 2.2, after 50h ph is constant at 6,42.

$$pH = -0{,}008\, x\, Age + 6{,}8218 \qquad \text{Equation 2.2}$$

with Age – time in hours since the inoculum was transferred into the pilot fermenter .

The agitation during phase I and II is constant at 300 rpm while in phase III it is manipulated between 300 to 1000 rpm to maintain pO_2 at a constant setpoint.

The pO_2 is also controlled by methanol feeding manipulation during the phase III after the reactor reaches the maximum oxygen transfer capacity (i.e. after the stirrer speed reaches 1000 rpm). The pO_2 set-point was set to 20% during this period.

The glycerol batch phase (phase I) takes approximately 30h to complete, after incubation. The beginning of phase II is signaled by the agitation rate, when it reaches the 440 rpm threshold value, which corresponds to a biomass of approximately 20 g/l. The duration of the glycerol fed-batch (phase II) was 12,5h. At the beginning of phase II, a small pulse of methanol (15g) was fed to the reactor to “promote” induction. This premature methanol feeding has the goal of facilitation the transition between glycerol and methanol phases.

The phase II used the following glycerol feeding formula

$$Q = Q_{0,Gly} \times e^{\mu.t} \qquad \text{Equation 2.3}$$

with Q = feeding rate (g/h); $Q_{0,Gly}$ = initial feeding rate, 30 g/h; μ_1 = desired sspecific growth rate, 0,16 h^{-1} for the initial 8,5h of this phase; and, for the last 4h, μ_2 = desired sspecific growth rate, 0,025 h^{-1}.

The duration of phase III, i.e. the methanol fed-batch phase, was approximately 100h. The methanol feeding was established by several optimizations in MATLAB™ (details are not shown here since this optimization is not within the scope of this thesis). The result was the following methanol feeding formula

$$y = 42{,}6036x + 0{,}0871x^2 + 0{,}0042x^3\ , \qquad \text{Equation 2.4}$$

y = Methanol Weight (kg) and x=Age (h).

All feeding streams (ammonium hydroxide for pH control, glycerol; and methanol) were weighted on-line using three weight scales with on-line data acquisition.

The UBICON, Universal Bio-Process Control System (Electronic System Design, Hanover, Germany), was used for data acquisition and process control, implementing all the control loops previously mentioned.

2.3. Biomass determination

The OD (optical density) of the broth was measured at 600nm and the samples were made in duplicate.

The wet cell weight (WCW) was determined by centrifugation of 13g of the fermentation broth at 15000 rpm for 10 minutes at 4ºC, and, also, the samples were made in duplicate.

2.4. ELISA-assay (for one plate)

The ELISA – Enzyme-Linked Immunosorbent Assay – is a technique for measuring the concentration of antibodies and antigens (Immunodectection Milipore®). The product concentration was obtained by this method.

There are several variations of ELISA that can be used for antibody assays:

- Direct ELISA: first the antigen is attached by passive adsorption to the well. After washing, enzyme-labeled antibodies are added and incubated for a period of time. After washing, the substrate that will provide color is added. The color appears and the reaction is stopped.

- Indirect ELISA: first antibodies react with the antigen and are attached by passive adsorption to the well. After washing, an anti-species antiserum labeled with enzyme are added and incubated for a period of time. After washing, the substrate that will provide color is added. The color appears and the reaction is stopped. This assay is widely used in diagnosis.

- Sandwich's ELISA: this assay is also divided into Direct Sandwich ELISA and Indirect Sandwich ELISA. The difference between these assays and the former ones is one extra layer of antibodies in the bottom of the well. First the antibody is

attached by passive adsorption to the well. After washing, the procedure is the same of the Direct ELISA or the Indirect ELISA. (Crowther, 2009)

In this work the assay applied was Sandwich Direct ELISA. The wells are coated with a purified antibody that will attach only the target protein. Then, an antibody linked with labeled molecule (the horseradish peroxidase – HRP) will attach the target protein. After that a substrate is added that is oxidized by HRP and will produce a color that will be analyzed by a spectrophotometer.

This method was performed using 96-well microplates (VWR). Add 100µl of protein A (50ul of protein A with 1g/L in 10ml of PBS (10mM pH7,4, Calbiochem)) in each well, incubate overnight at 4ºC with moist environment or for one hour at 37ºC; wash three times with Tween-PBS (pH 7,4 ,Calbiochem) and dry; add 200µl of blocking solution (Tween-PBS with 5%(w/v) skim milk (Merck)) in each well and incubate for one hour at 37ºC with moist environment; wash three times with Tween-PBS and dry. In the next step standard (purified scFv) – positive control, PBS – negative control, and samples were applied after dilution with skim milk and duplicate – incubate for 1 hour at room temperature; wash three times with Tween-PBS. Incubate the plate with 100µl Anti-human IgG-HRP (Sigma) (20µl in 10ml of blocking solution) in each well and incubate for 30 minutes at 37ºC, wash three times with Tween-PBS. Add 100µl of coloring solution (OPD peroxidase, 1 tablet in 20ml of water,Sigma), protect form light during this process. The reaction was stopped with 50µl of Sulfuric acid (2,25N, Merck). The absorbance was measured at 450-490nm using a microplate reader (Izasa). The concentration was calculated by linear regression.

2.5. Purification

The purification of the antibody fragment was done by two-steps chromatography: first affinity chromatography with protein A as a capture step; then a size-exclusion chromatography as second step to separate the monomer and dimmer forms.

The affinity chromatography with protein A is based in the proprieties of this protein. Protein A is a single polypeptide chain and the affinity of this chromatography resides in the hydrophobicity of the binding site. When the conditions are alkaline the residues at the binding site are uncharged and hydrophobic, so the interaction between protein A and the antibody is strong. While pH is shifted to acidic values the interaction

becomes week, the residues are charged and there is repulsion between protein A and the antibody (Bailon, et al., 2000).

The size-exclusion chromatography is a liquid chromatography that provides separation of macromolecules by molecular size. The column has a packing material of porous particles which form the stationary phase. The liquid mobile phase begins to pass through the column, the smaller macromolecules are able to penetrate into the pores of the stationary phase and the larger macromolecules can't penetrate into the pores so they flow more rapidly. In the end, there are two distant chromatographic bands with two different molecular sizes (Wu, 2004).

Fermentation broth supernatant, collected by centrifugation at 13,000g for 15 min, after pH adjustment to 7.4 with 1 M NaOH, was loaded at a linear velocity of 300 cm h^{-1} onto a Streamline 50 column (50 mm diameter, Amersham Biosciences, Piscataway, NJ) filled with 300 mL of Protein A. The column was equilibrated with PBS, 10% glycerol at pH 7.4. The loaded column was then washed with the same buffer until a stable baseline was obtained and eluted with 0.1 M citrate, 10% glycerol at pH 3.5 at 150 cm h^{-1} followed by immediate neutralization with 1 M Tris base. Eluate (10 mL) at a concentration of approximately 1.5 mg mL^{-1} was then loaded onto a gel filtration HiLoad Superdex 75 26/60 column (Amersham Biosciences) and equilibrated with PBS and 10% glycerol at pH 7.4. (Cunha, *et al*., 2004)

2.6. BCA Protein Assay Kit

This is a colorimetric detection assay for quantification of total protein after purification.

This method is based on reduction of Cu^{2+} to Cu^{+} by protein in an alkaline medium with a selective colorimetric detection of the cuprous cation (Cu^{+}).

The protein identification is made by the transition of a colorless solution to an intense purple color with an absorbance maximum at 562nm. It was used a kit for this assay with the protein concentration range 20-2000µg/ml and with BSA (bovine serum albumin) as a reference protein.

The BCA Protein Assay Kit (Pierce) was used according to the manufacturer's instructions using a microplate procedure. In the end of the procedure the volume was

50 µl of sample plus 200 µl of BCA reagent. The assay was conducted at 37ºC for 30 min and the absorbance was measured at λ = 562 nm.

The BSA was the reference protein and the dilution curve consisted of eight points in duplicate with the concentration between 0,1 – 2 mg/ml. The scFv was also made in duplicate and the concentration was calculated by linear regression.

2.7. MALDI

The MALDI technique is a mass spectrometer technique to identify, verify and quantify proteins isolated from natural sources or recombinant proteins, metabolites, oligonucleotides, drugs, peptides, polymers, and organic compounds. Because of the large amount of compounds, this technique is applied in many sectors, like pharmaceuticals analysis, environmental analysis and biomolecular characterization (Hillenkamp *et al*., 2007).

Very briefly, the MALDI spectyrometer works as follows. The samples are mixed with the matriz and dried on a plate and this plate is put inside of the spectrometer. Inside of the spectrometer, there are three main components: the ion source that will ionize the molecules of the sample; the mass analyzer, which separates the ions based in the ratio mass/charge (m/z), and that can be Time-of-flight (TOF), quadruple or an ion trap; and the detector, which registers the ions that arrive and measure the ratio (m/z). Finally a computer registers and processes the data providing the mass spectrum (AguardLab; Young).

The MALDI technique has several advantages over other mass spectrometry techniques, it can analyze intact biomolecules and synthetic polymers; it can analyze a large variety of compounds, with masses over 300kDa; it can be relatively tolerant to buffers and salts; it has high sensitivity and is fast (AguardLab).

Substitution of the glycerol and concentration of the sample

200 µL of PBS buffer was placed in the Amicon Ultra-4Centrifugal Filter to balance the membrane and centrifuged (4000rpm for 5min. at 7ºC). 200µL of the protein (the scFv) with the concentration of 2,955 mg/ml was mixed with 1800 µL of PBS buffer, agitated with micropipette and centrifuged (4000rpm at 7ºC) until 200 µL were reached. More PBS buffer is added if necessary. The concentration of product over time is measured using the BCA Protein Assay Kit.

10 µl of the received protein solution were reduced, alkylated and digested in-solution by trypsin in the presence of 1M urea (final volume 80 µl). 10 µl tryptic and PGNase F digests were desalted, concentrated and eluted using a R2 micro-column (RP-C18 equivalent). Peptides were eluted directly onto a MALDI plate with α-ciano-4-hidroxycinamic acid (CHCA) (5 mg/ml) in 50% acetonitrile, 5% formic acid. Mass spectra of the peptide mixtures were acquired in the positive reflectron MS and MS/MS modes using a MALDI-TOF/TOF MS (4800plus MALDI TOF/TOF analyzer). The acquired MS and MS/MS spectra were analyzed in combined mode using Mascot search engine and NCBI database without taxonomic restrictions with a peptide mass tolerance of 50 ppm[1].

2.8. Bioinformatic methods

These methods are all available in http://www.expasy.org/tools/ (Gasteiger, *et al*., 2005) with an HTML interface.

2.8.1. Peptide Cutter

This tool predicts cleavage sites caused by several proteases or chemicals for a given protein sequence.

The protein sequence has to be entered in the form of Swiss-Prot / TrEMBL accession number (UniProtKB); or a sequence in FASTA format (sequence of one-letter aminoacid code).

Once the protein sequence is coded, it is possible to select one or multiple enzymes or chemicals provided in the list.

It is possible to display the final map of cleavage sites in different ways: a table of sites ordered alphabetically by enzyme or chemical name; or a table of sites ordered alphabetically by aminoacid number. In the latter way, it is possible to choose the exact time of cut that the enzyme or chemical do or a range of cleavages (Gasteiger, *et al*., 2005).

[1] Data provided/obtained by the Mass Spectrometry Laboratory, Analytical Services Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa.

Protein sequence of AP39 was introduced and it was selected Trypsin in "only the following selection of enzymes and chemicals".

2.8.2. Peptide Mass

The Peptide Mass tool has the purpose to help in peptide-mapping experiments, and help with the interpretation of mass-spectrometry data.

The first step is entering the protein sequence. The protein sequence can be coded in the same forms accepted by the PeptideCutter.

Then it is possible to choose several modifications that can be performed to the protein; the adduct that was used in the experiment; and if the mass is displayed by average or monoisotopic.

Next it is possible to choose the enzyme used for protein cleavage, if there are missed cleavages and the range of the peptide mass to consider (Gasteiger, *et al*., 2005).

Protein sequence of AP39 was introduced, it was selected: the peptide masses are $[M+H]^+$ and monoisotopic; the enzyme Trypsin; allow for 3 missed cleavages.

2.8.3. GlycoMod

The GlycoMod tool finds all possible composition of glycan structures from the experimentally determined mass.

The first step is entering the peaks list of experimental masses (either average or monoisotopic measured masses) and the mass tolerance in Dalton or ppm.

The second step is choosing the ion mode and the adducts that were used during the MALDI experiment.

Next we need to choose between: N-linked oligosaccharides in the form of free oligosaccharides (with PNGase F enzyme), in the form of released oligosaccharides (with EndoH or EndoF enzymes), in the form of reduced oligosaccharides, in the form of derivatised oligosaccharides, or in the form of glycopeptides (the glycans are still attached in the peptide); or choose O-linked oligosaccharides in the form of free oligosaccharides, in the form of reduced oligosaccharides, in the form of derivatised oligosaccharides, or in the form of glycopeptides.

The fourth input is the protein sequence with the cleavage enzyme and the possible modifications that the protein can have.

The last input is the monosaccharide residue (if known). It is possible to choose the monosaccharides that exist in each case, and if they are underivatised, permethylated or peracetylated (Gasteiger *et al*., 2005).

Experimental masses and protein sequence of AP39 were introduced. It was selected: the peptide masses – monoisotopic; the ion mode and adducts – $[M+H]^+$; the enzyme Trypsin, allow for 3 missed cleavages; form of O-linked oligosaccharides – Glycopeptides (only those containing S or T will be used); monosaccharide residues are underivatised; and it was chosen Hexose Man (selected as yes) and all the others were eliminated (selected as no).

2.9. Glycosylation mathematical models

Glycosylation mathematical models are based on the reactions that occur within a plug flow reactor (Figure 2.1) (Shelikoff, *et al*., 1996). For the Endoplasmatic Reticulum (E.R.) there are the inflows of protein (C_P) and mannoses (C_M); and the outflows of protein (C_P), mannoses (C_M), protein with one mannose (C_{PM}) and protein with two mannoses (C_{PM_2}). The inflows of the *Golgi* are the outflows of the E.R. and the outflows are protein (C_P), mannoses (C_M), protein with one mannose (C_{PM}), protein with two mannoses (C_{PM_2}), protein with three mannoses (C_{PM_3}), protein with four mannoses (C_{PM_4}) and protein with five mannoses (C_{PM_5}). The mathematical models are based on the Equation 2.5 to Equation 2.8, where C_i is the concentration of the several compounds, the C_{M_0} is the initial concentration of mannoses and t_r is the residence time.

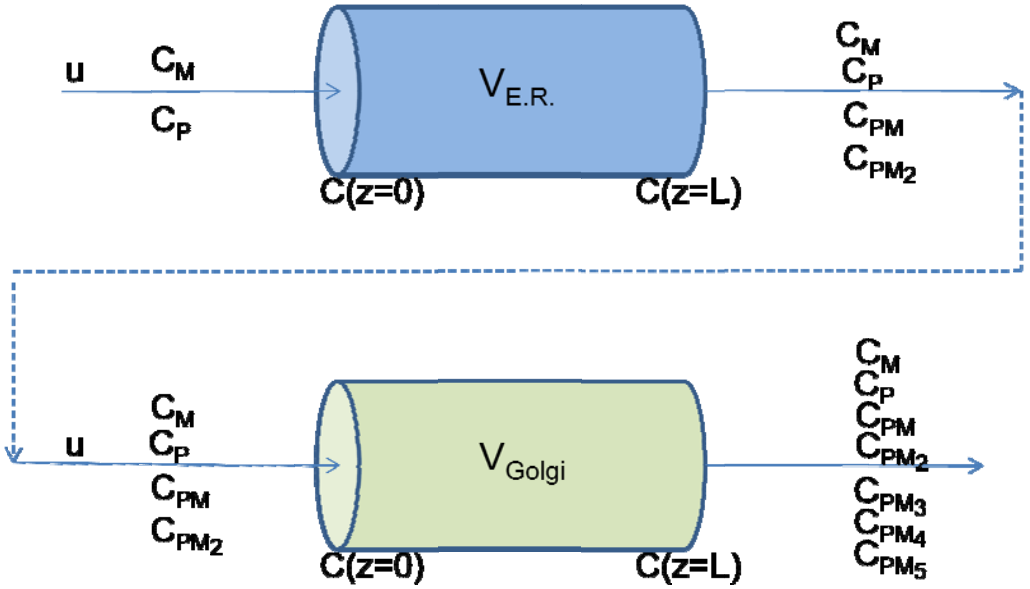

Figure 2.1 – Glycosylation model in *P pastoris:* analogy with a Plug Flow Reactor (PFR) system

The material balances of each compound in the PFR are

$$\frac{d\,C}{d\,t} = \boldsymbol{A}\,r \qquad \text{Equation 2.5}$$

with $C = [C_M, C_P, C_{PM}, C_{PM_2}, C_{PM_3}, C_{PM_4}, C_{PM_5}]^T$, A = a stoichiometric matrix (see below) and $r = [r_1, r_2, r_3, r_4, r_5]^T$.

$$\mathrm{A}= \begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \qquad \text{Equation 2.6}$$

The material balances can be transformed into a dimensionless form as follows:

$$\frac{dC^*}{d\,t^*} = A\,r^* \qquad \text{Equation 2.7}$$

$$C^* = \frac{C}{C_{M_0}} \qquad \text{Equation 2.8}$$

$$\boldsymbol{t^*} = \frac{\boldsymbol{t}}{\boldsymbol{t_r}} \qquad \text{Equation 2.9}$$

$$\boldsymbol{r^*} = \boldsymbol{r}\frac{\boldsymbol{t_r}}{\boldsymbol{C_{M_0}}} \qquad \text{Equation 2.10}$$

In the equations below superscript * is omitted for simplicity. All the results shown in this thesis used the dimensionless from of material balances

Applying Equation. 2.6 to the ER results:

$$\frac{d}{dt}\begin{bmatrix} C_M \\ C_P \\ C_{PM} \\ C_{PM_2} \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \qquad \text{Equation 2.11}$$

Applying Equation. 2.6 to the *Golgi* results:

$$\frac{d}{dt}\begin{bmatrix} \boldsymbol{C_M} \\ \boldsymbol{C_P} \\ \boldsymbol{C_{PM}} \\ \boldsymbol{C_{PM_2}} \\ \boldsymbol{C_{PM_3}} \\ \boldsymbol{C_{PM_4}} \\ \boldsymbol{C_{PM_5}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{-1} & \boldsymbol{-1} & \boldsymbol{-1} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{1} & \boldsymbol{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{-1} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{r_3} \\ \boldsymbol{r_4} \\ \boldsymbol{r_5} \end{bmatrix} \qquad \text{Equation 2.12}$$

The model above assumes that the glycosylation of O-linked glycans is a deterministic reaction process, i. e. there is no randomness involved. However for very low mannose concentrations it is likely that the reaction of a mannose molecule with several alternative molecules of P, PM_1, PM_2, PM_3, PM_4 and PM_5 is a random event and thus a stochastic model formulation is required. This requirement is further reinforced by the strong possibility of mannose being the limiting factor for glycosylation.

The *Tau leaping* method (Gillespie, 2005) was adopted for stochastic simulation. According to this method, the number of times that a reaction j occurs in a time interval $[t\,, t+\tau]$is given by the Poisson distribution,

$$\boldsymbol{P}\left(\boldsymbol{n_{r,j}}, \boldsymbol{\lambda}\right) = \frac{\boldsymbol{\lambda_j^{n_{r,j}}}\, \boldsymbol{e^{-\lambda_j}}}{\boldsymbol{n_{r,j}!}} \qquad \text{Equation 2.13}$$

with mean, λ_j given by

$$\boldsymbol{\lambda} = \boldsymbol{r_j}\, \boldsymbol{N_{MAN,0}}\, \boldsymbol{\tau} \qquad \text{Equation 2.14}$$

Notice that r_j is the dimensionless reaction rate value, relative to the initial mannose concentration. In stochastic simulation the total number of molecules is determinant for the probability of a reaction to occur. Thus r_j is here multiplied by the total number of mannose molecules, $N_{MAN,0}$, in order to calculate the number of mannose molecules that react per unit residence time.

Once the number of times reactions occur is evaluated for j=1,…,5, using a Poisson random number generator according to Equation. 2.13 and Equation 2.14 (the *poissrnd* MATLAB function was adopted), then the concentrations of compounds at time $t + \tau$ are evaluated as

$$\boldsymbol{C}\,(t+\tau) = \boldsymbol{C}(t) + \boldsymbol{A}.\boldsymbol{n} \qquad \text{Equation 2.15}$$

with

$$\boldsymbol{n_r} = \left[\boldsymbol{n_{r,1}}, \boldsymbol{n_{r,2}}, \boldsymbol{n_{r,3}}, \boldsymbol{n_{r,4}}, \boldsymbol{n_{r,5}}\right]^{\boldsymbol{T}} \qquad \text{Equation 2.16}$$

Both mathematical models have the same initial conditions: concentrations at the entrance of the PFR ($C_M/C_{M_0} = 1$, C_{P_0}/C_{M_0} varied, $C_{PM_1}/C_{M_0} = C_{PM_2}/C_{M_0} = C_{PM_3}/C_{M_0} = C_{PM_4}/C_{M_0} = C_{PM_5}/C_{M_0} = 0$). For the stochastic model, it is necessary to define the total number of mannose molecules at the entrance of the PFR, $N_{MAN,0}$, which was varied to assess the effect of total number of molecules in the glycosylation dynamics (MATLAB code in Annex H)

2.10. Glycoprotein Detection Kit

The glycoprotein detection kit is a simple and easy system to detect glycoproteins on a SDS-PAGE or a Western blotting membrane. This kit is based on the modification of Periodic Acid-Schiff (PAS) method. First there is a reaction involving the oxidation of carbohydrates by Periodic acid, which oxidizes 1,2-glycol groups to aldehydes; and the second step is the staining with Schiff's reagent, the aldehydes react with this reagent to form a magenta or pink colored band (Moller *et al*., 1996; Yan *et al*., 2008; Deepak *et al*., 2003). In order to this method to be reliable it is necessary several micrograms of proteins for the detection (Moller, *et al*., 1996).

Before applying the glycoprotein detection kit it was necessary to make a polyacrylamide gel electrophoresis (PAGE). The system applied was the NuPAGE electrophoresis system (Invitrogen), it was used 4-12% Bis-Tris Gel. The sampler

buffer (was used 2x concentrate – 0,08M Tris-HCl, ph 6,8 (Merck); 2% (w/v) SDS (Merck); 0,1M DDT (Calbiochem); 10% (v/v) glycerol (Vaz Pereira); 0,001% Bromophenol blue (Sigma)) where the respective protein was dissolved and was denaturated at 100ºC for 3 min. The running buffer was MES 20x (Invitrogen). The gel was loaded and run at 200V (constant) for approximately 45min.

Next it was applied the glycoprotein detection kit according to the manufacturer's instructions as follows: first immerge the gel in the fixing solution (50% (v/v) methanol) gently agitate for one hour; wash with pure water, gently agitate for twenty minutes, repeat the process; immerge the gel in Oxidation solution (this solution was provided by the kit) gently agitate for one hour; repeat the process of washing; immerge the gel in Staining solution (provided by the kit), gently agitate for 1-2 hours or until bands turn to magenta; replace this solution for the Reduction solution, gently agitate for two hours; wash with pure water, repeat two or three times.

After these steps, the glycoprotein was completely colored and the gel was scanned. The gel was colored with 20 ml of SimpleBlue^{TM} (Invitrogen) for 1 hour at room temperature, for identifying other proteins in the gel. The gel was scanned one more time after this staining.

3. Results

3.1. Fermentation

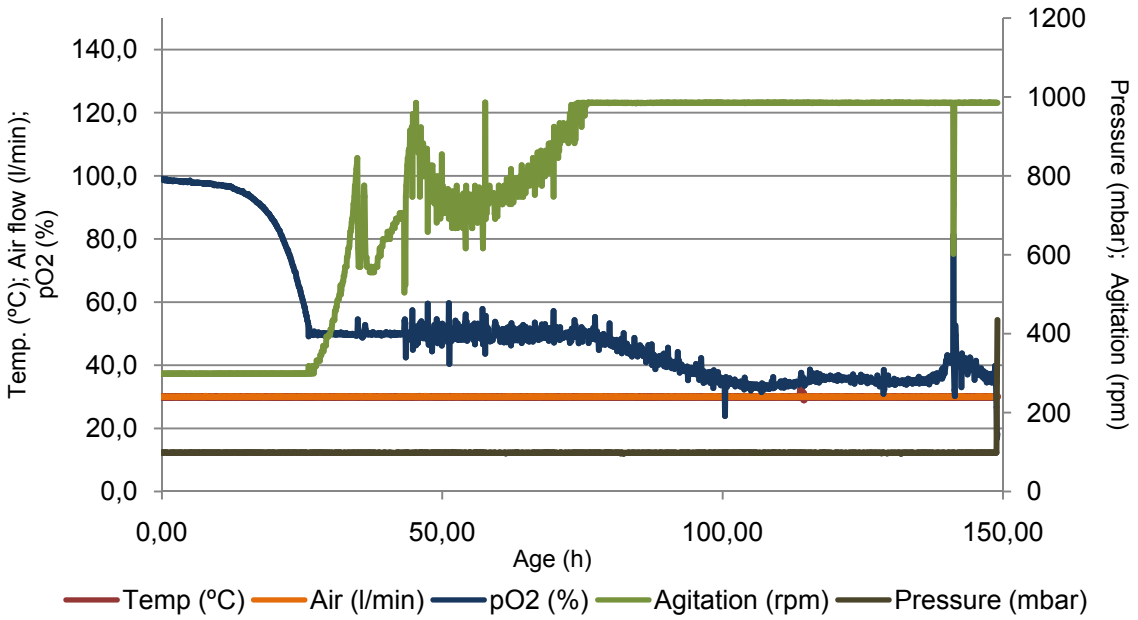

Figure 3.1 – Fermentation variables over time. It should be noticed that the temperature and the airflow rate are overlapped.

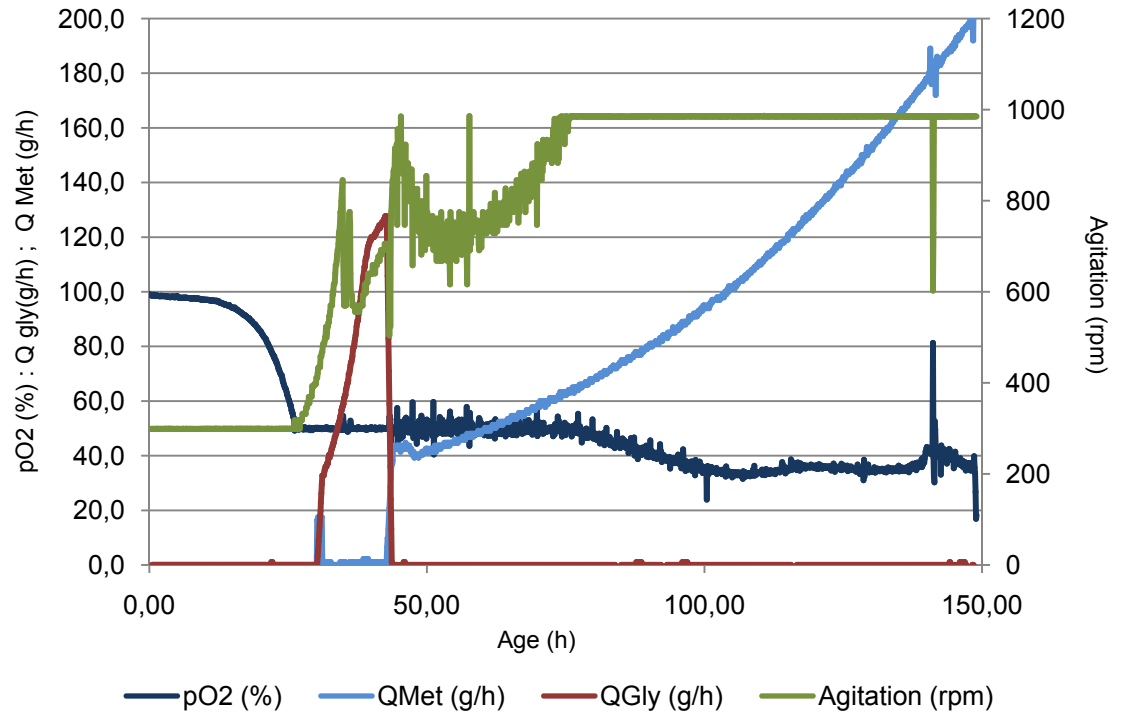

Figure 3.2 – Carbon sources feeding, glycerol and methanol; pO_2 and agitation over fermentation time.

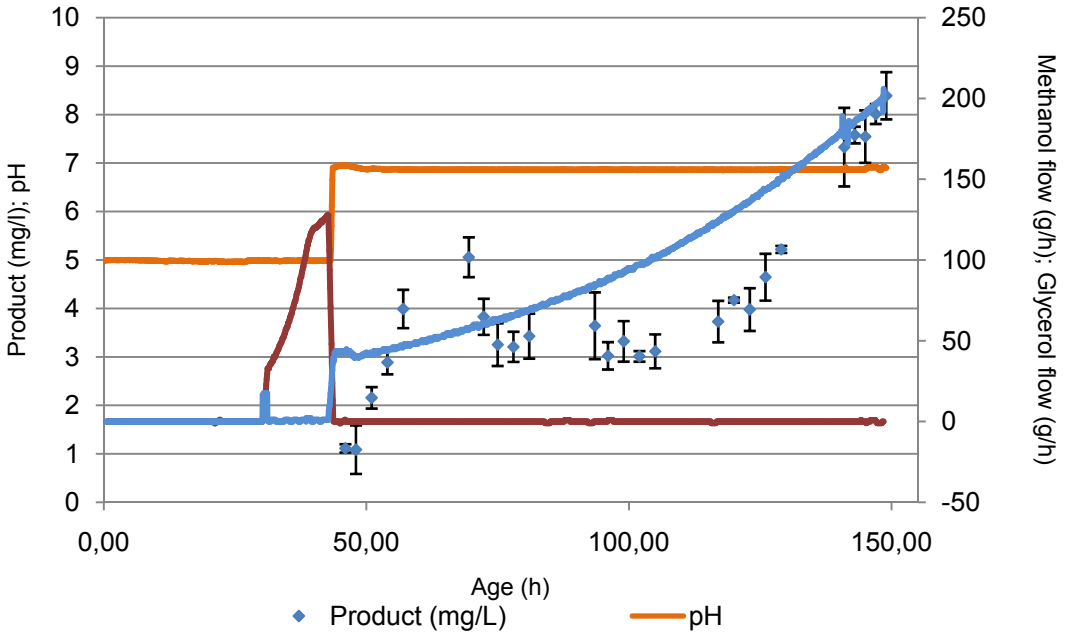

Figure 3.3 – Feeding rates of carbon sources, glycerol and methanol; ph and product concentration over time.

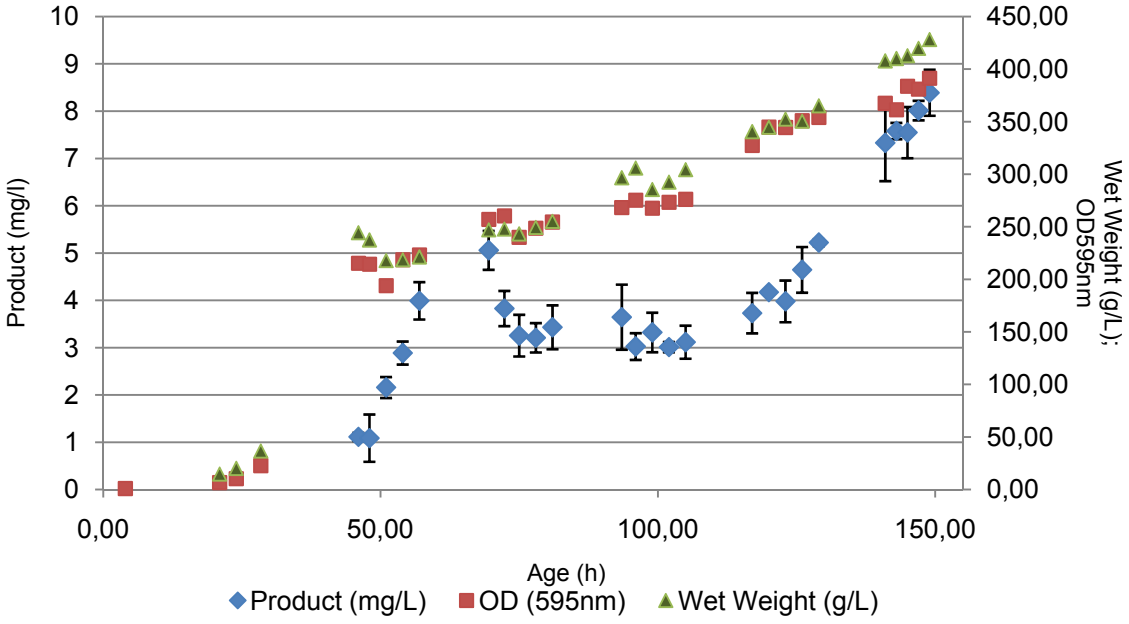

Figure 3.4 - Results of the biomass determination with OD (595nm) and wet weight and product concentration with ELISA assay (Annex B), over fermentation time.

3.2. BCA Protein Assay kit

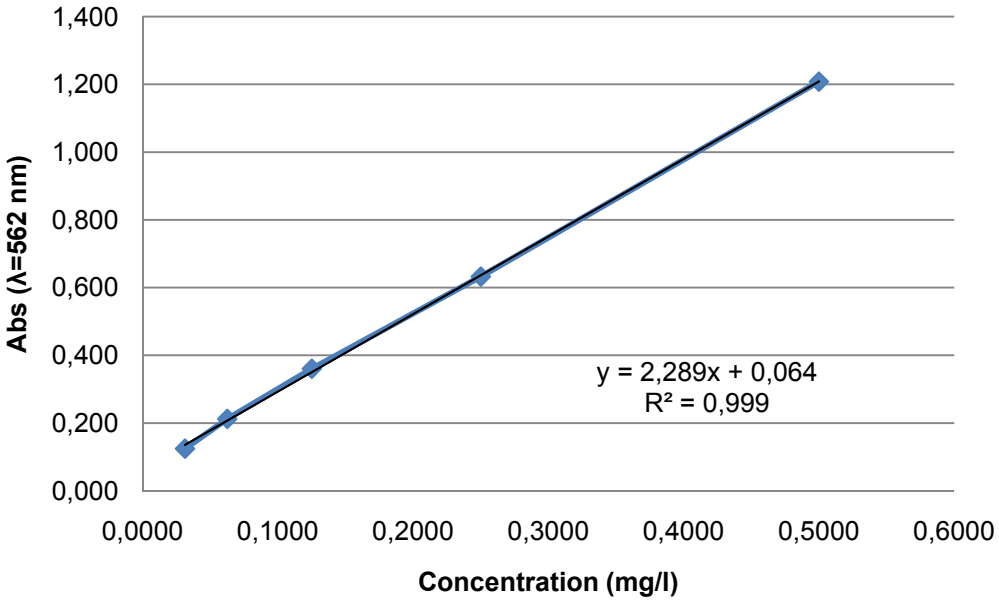

Figure 3.5 – Calibration curve for BCA Protein Assay Kit with BSA as standard protein (data in Annex A).

Table 3.1 – Concentrations of the product after purification and after using the Amicon Ultra-4Centrifugal Filter with their dilution factor.

	Product after purification		**Product after concentration and substitution of glycerol**	
Dilution factor	**Absorvance average**	**Concentration (mg/ml)**	**Absorvance average**	**Concentration (mg/ml)**
2	2,4100	1,9912	2,3760	1,9615
4	1,5820	2,5354	1,9370	3,1558
8	0,9445	2,8428	1,1480	3,5540
16	0,5630	3,0189	0,6990	3,9695
32	0,3545	3,1230	0,4460	4,4022
64	0,2325	2,8350	0,2810	4,1910
128	0,1500	1,0567	0,1770	2,5665
256	0,1155	-1,7451	0,1280	-0,3471

3.3. MALDI analysis

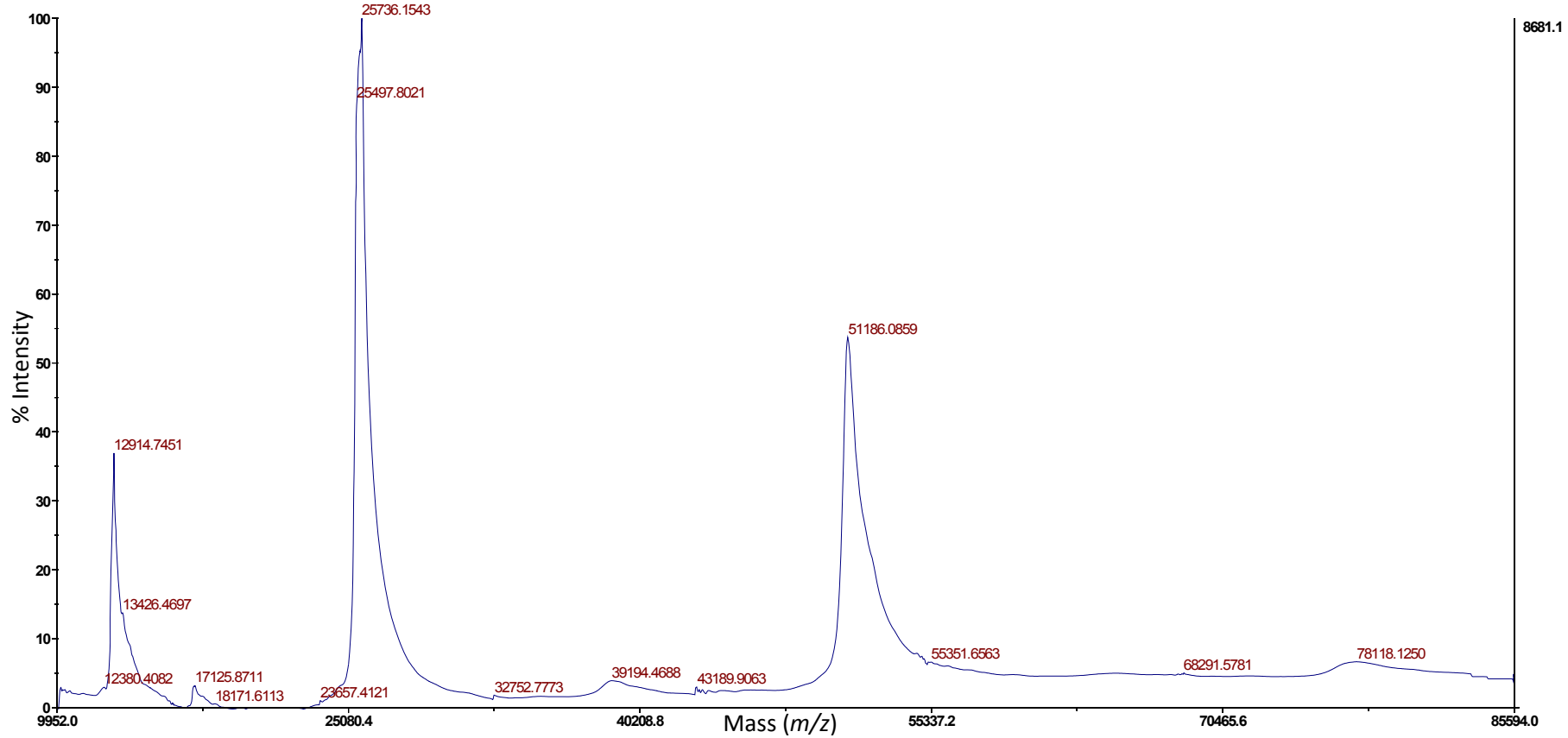

Figure 3.6 – MALDI spectrum of intact protein[2]. Zoom in of the major peaks in Annex C.

[2] Data provided/obtained by the Mass Spectrometry Laboratory, Analytical Services Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa.

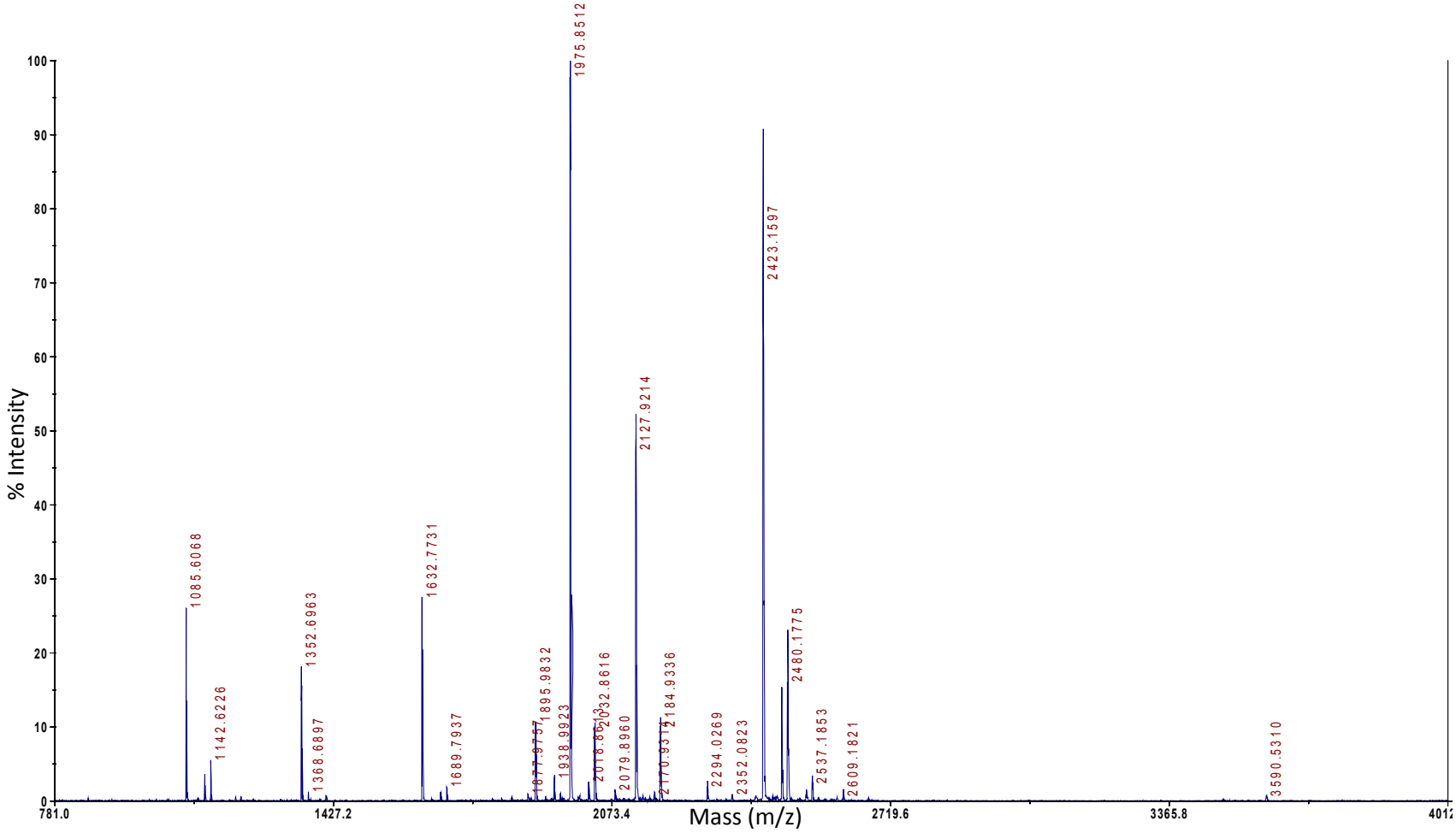

Figure 3.7 – MALDI spectrum of the protein digested by trypsin[3]. List of the peaks of this spectrum in Annex D, Table 7.5.

[3] Data provided/obtained by the Mass Spectrometry Laboratory, Analytical Services Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa".

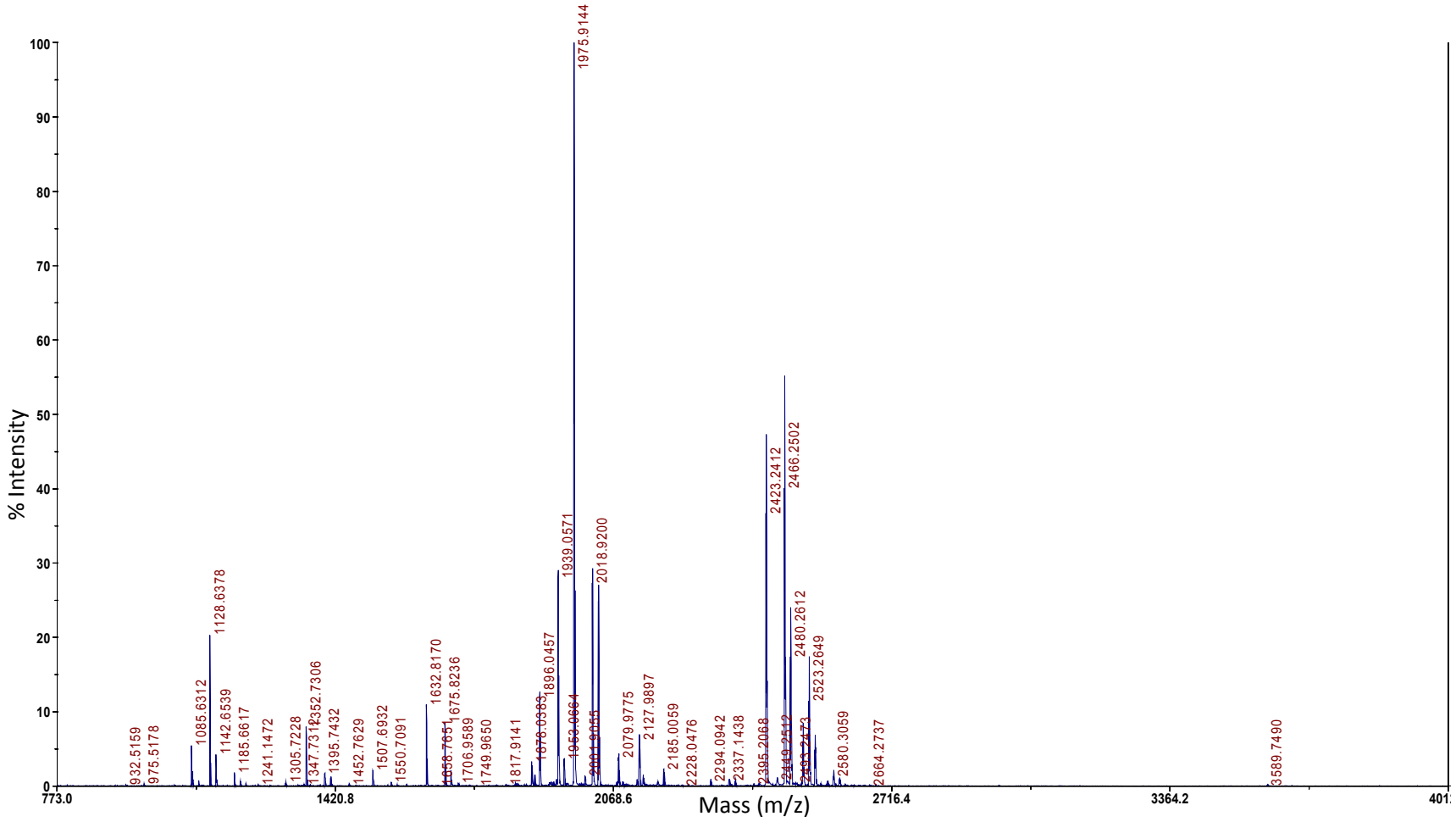

Figure 3.8 – MALDI spectrum of the protein cleaved by trypsin and PNGase F[4]. List of the peaks of this spectrum in Annex D, Table 7.6Table 7.5.

[4] Data provided/obtained by the Mass Spectrometry Laboratory, Analytical Services Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa".

```
10         20         30         40         50         60
EVQLLESGGG LVQPGGSLRL SCAASGFTFS SFSMSWVRQA PGKGLEWVSS ISGSSGTTYY

70         80         90        100        110        120
ADSVKGRFTI SRDNSKNTLY LQMNSLRAED TAVYYCAKPF PYFDYWGQGT LVTVSSGDGS

130        140       150        160        170        180
SGGSGGASTG EIVLTQSPGT LSLSPGERAT LSCRASQSVS SSFLAWYQQK PGQAPRLLIY

190        200        210        220        230        240
YASSRATGIP DRFSGSGSGT DFTLTISRLE PEDFAVYYCQ QTGRIPPTFG QGTKVEIKGG GCA
```

Figure 3.9 – Aminoacid sequence of AP039.

3.4. Bioinformatic methods

Table 3.2 - List of the peptide sequence and their masses (from PeptideCutter and PeptideMass).

Mass (m/z)	Position	Peptide sequence
6177,8821	88-148	AEDTAVYYCAKPFPYFDYWGQGTLVTVSSGDGSSGGSGGA STGEIVLTQSPGTLSLSPGER
2423,2102	155-176	ASQSVSSSFLAWYQQKPGQAPR
2294,0823	44-65	GLEWVSSISGSSGTTYYADSVK
2070,9412	20-38	LSCAASGFTFSSFSMSWVR
1918.8640	209-224	LEPEDFAVYYCQQTGR
1896,0185	1-19	EVQLLESGGGLVQPGGSLR
1632,7864	193-208	FSGSGSGTDFTLTISR
1352,6991	77-87	NTLYLQMNSLR
1085,5989	177-185	LLIYYASSR
1045.5676	225-234	IPPTFGQGTK
729,3889	186-192	ATGIPDR
650,3290	149-154	ATLSCR
623,3511	68-72	FTISR
500,2827	39-43	QAPGK
462,2074	73-76	DNSK

Table 3.3 – Determination of possible glycoforms by GlycoMod.

Position	Sequence	Experimental mass (m/z)	Glycoform mass (m/z)	Δmass (Dalton)	Peptide mass	Structure
77-87	NTLYLQMNSLR	2001,9053	648,211	0,995	1351,692	$(Man)_4$
177-185	LLIYYASSR	1409,7137	324,106	0,009	1084,592	$(Man)_2$
177-186	LLIYYASSR	1895,9832	810,264	0,121	1084,592	$(Man)_5$
209-224	LEPEDFAVYYCQQTGR	2079,9773	162,053	-0,939	1917,857	$(Man)_1$
209-225	LEPEDFAVYYCQQTGR	2566,283	648,211	-0,791	1917,857	$(Man)_4$
225-235	IPPTFGQGTK	1368,6901	324,106	-0,983	1044,56	$(Man)_2$
68-76	FTISRDNSK (1 missed cleavage)	1877,9755	810,264	0,164	1877,812	$(Man)_5$
66-76	GRFTISRDNSK (2 missed cleavage)	2089,9597	810,264	-0,974	1279,663	$(Man)_5$
66-76	GRFTISRDNSK (2 missed cleavage)	1927,9136	648,211	-0,967	1279,663	$(Man)_4$
225-238	IPPTFGQGTKVEIK (1 missed cleavage)	2001,9053	486,158	0,89	1513,85	$(Man)_3$
225-243	IPPTFGQGTKVEIKGGGCA (2 missed cleavage)	2184,9336	324,106	0,859	1858,961	$(Man)_2$

3.5. Glycosylation mathematical models

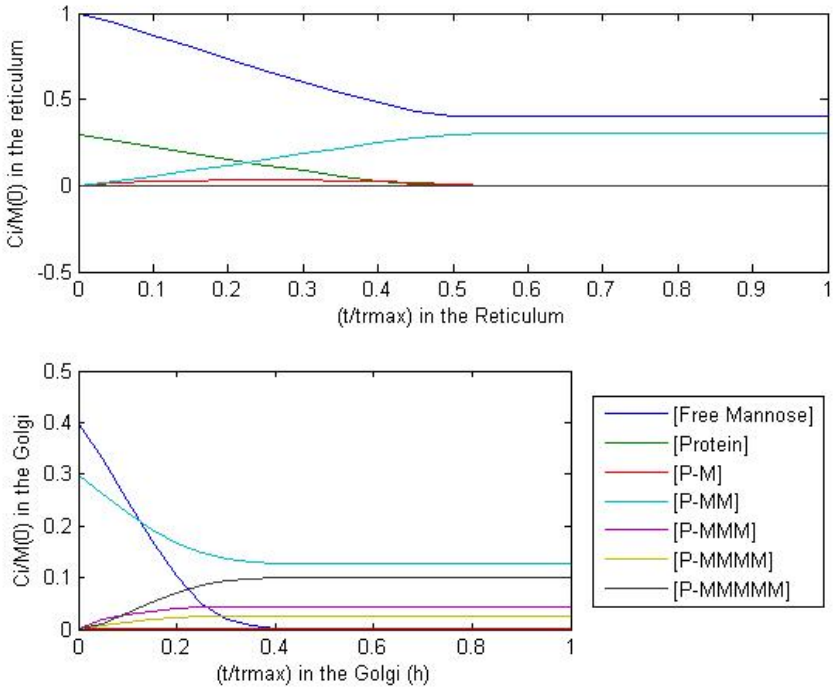

Figure 3.10 – Deterministic model results with a ratio of C_P/C_{M_0} of 0,3; and 100 mannose molecules at the entrance of PFR.

Table 3.4 – Composition out of the *Golgi* in a deterministic model with a ratio of C_P/C_{M_0} of 0,3; and 100 mannose molecules at the entrance of PFR.

	Outlow	**Percentage of C_{P_0}**
C_M/C_{M_0}	0	0
C_P/C_{M_0}	0	0
C_{PM_1}/C_{M_0}	0	0
C_{PM_2}/C_{M_0}	0,13	42,65
C_{PM_3}/C_{M_0}	0,045	14,94
C_{PM_4}/C_{M_0}	0,026	8,83
C_{PM_5}/C_{M_0}	0,10	33,58

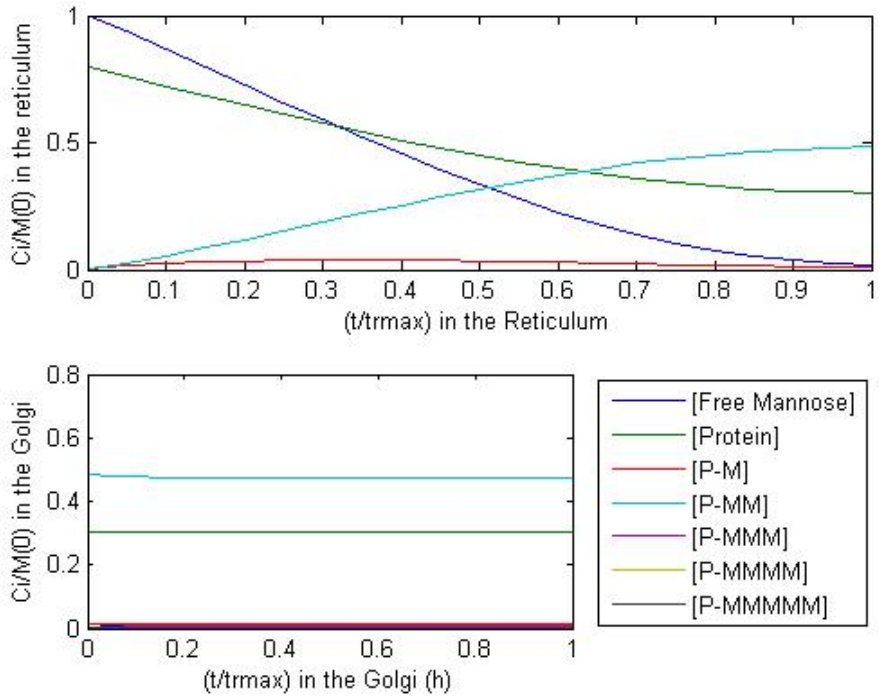

Figure 3.11 – Deterministic model results with a ratio of C_P/C_{M_0} of 0,8; and 100 mannose molecules at the entrance of PFR.

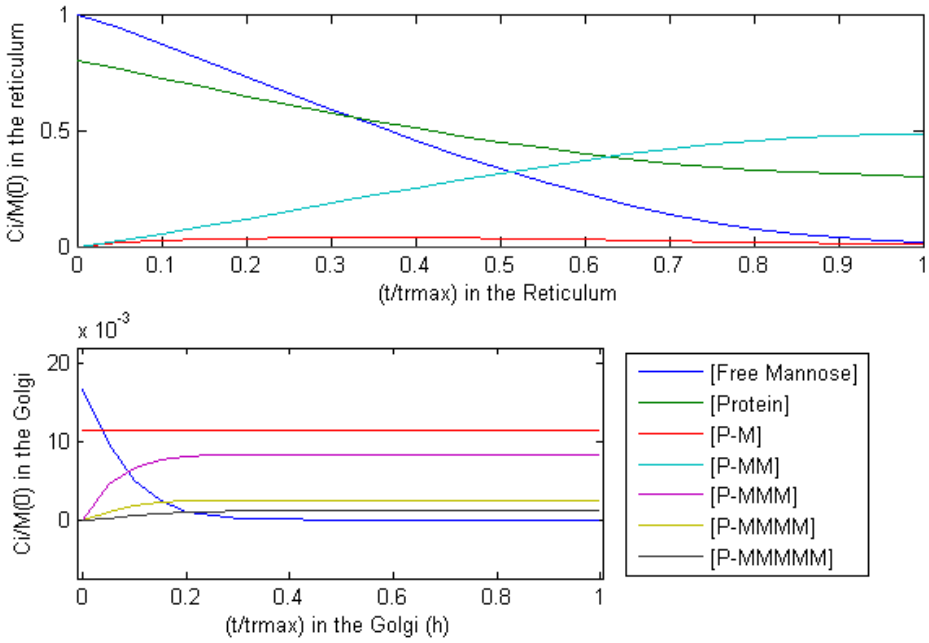

Figure 3.12 – Zoom in the graphic of *Golgi* of Figure 3.11.

Table 3.5 - Composition out of the *Golgi* in a deterministic model with a ratio of C_P/C_{M_0} of 0,8 ; and 100 mannose molecules at the entrance of PFR.

	Outlow	**Percentage of C_{P_0}**
$\mathbf{C_M/C_{M_0}}$	0	0,000012
$\mathbf{C_P/C_{M_0}}$	0,30	37,83
$\mathbf{C_{PM_1}/C_{M_0}}$	0,012	1,44
$\mathbf{C_{PM_2}/C_{M_0}}$	0,47	59,22
$\mathbf{C_{PM_3}/C_{M_0}}$	0,0084	1,05
$\mathbf{C_{PM_4}/C_{M_0}}$	0,0025	0,32
$\mathbf{C_{PM_5}/C_{M_0}}$	0,0011	0,14

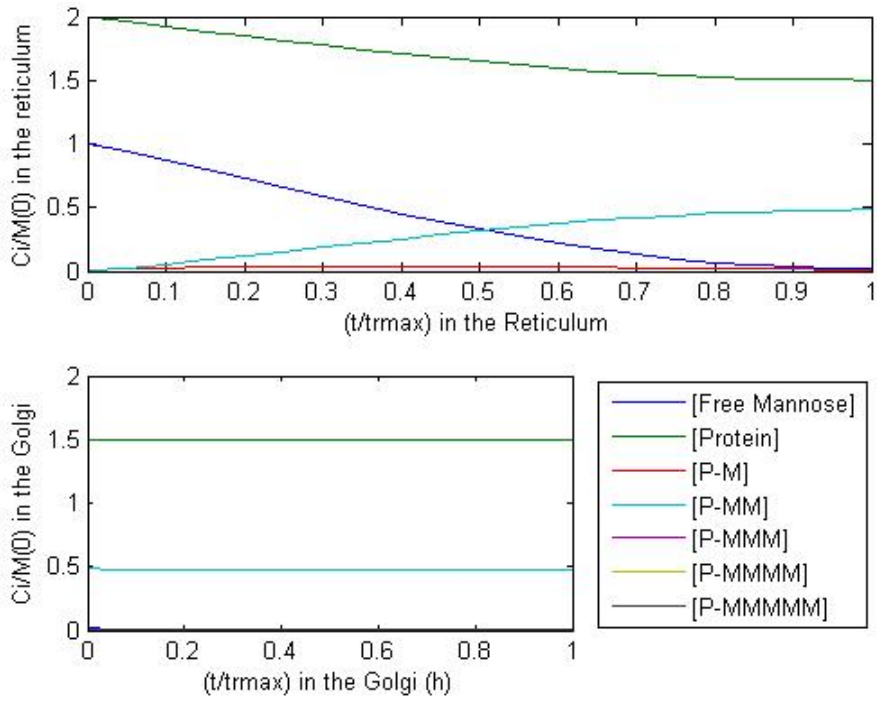

Figure 3.13 - Deterministic model results with a ratio of C_P/C_{M_0} of 2; and 100 mannose molecules at the entrance of PFR.

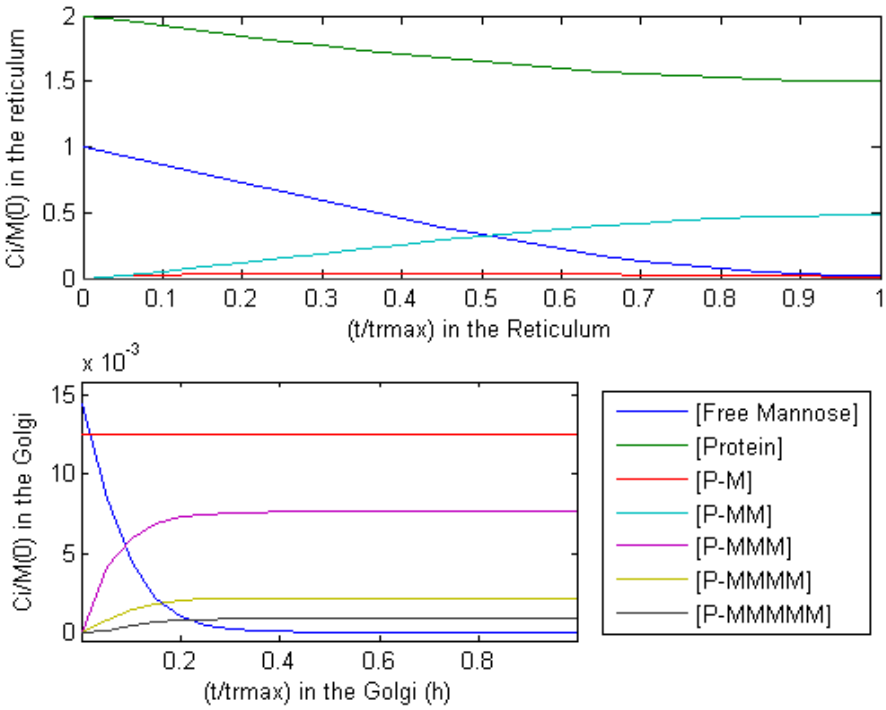

Figure 3.14 – Zoom in the graphic of *Golgi* of Figure 3.13.

Table 3.6 – Composition out of the *Golgi* in a deterministic model with a ratio of C_P/C_{M_0} of 2 and 100 mannose molecules at the entrance of PFR.

	Outlow	**Percentage of C_{P_0}**
C_M/C_{M_0}	0	0
C_P/C_{M_0}	1,50	75,05
C_{PM_1}/C_{M_0}	0,013	0,63
C_{PM_2}/C_{M_0}	0,48	23,79
C_{PM_3}/C_{M_0}	0,0076	0,38
C_{PM_4}/C_{M_0}	0,0022	0,11
C_{PM_5}/C_{M_0}	0,00083	0,043

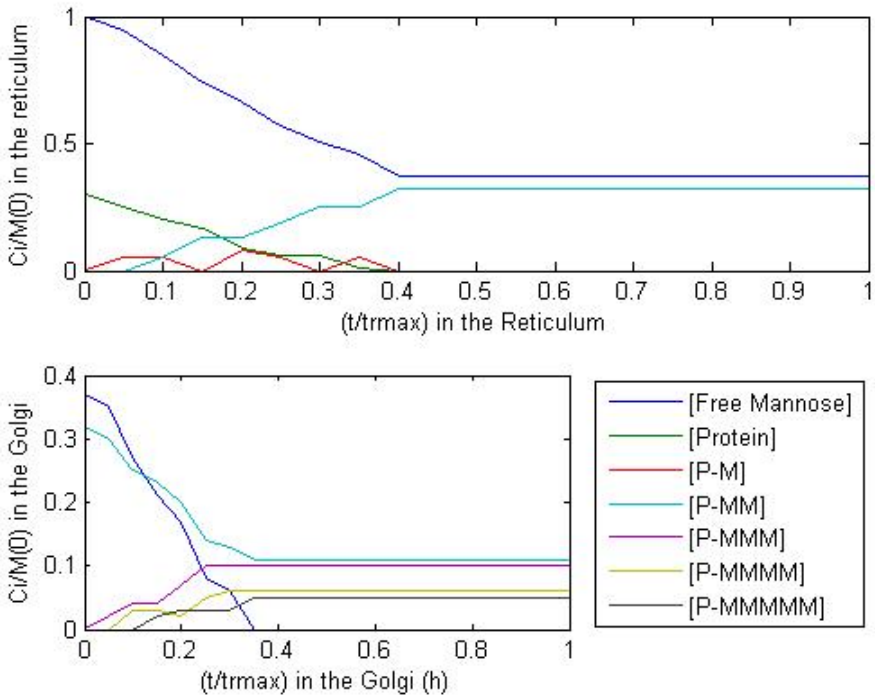

Figure 3.15 – Stochastic model results with a ratio of C_P/C_{M_0} of 0,3 and 100 mannose molecules at the entrance of PFR.

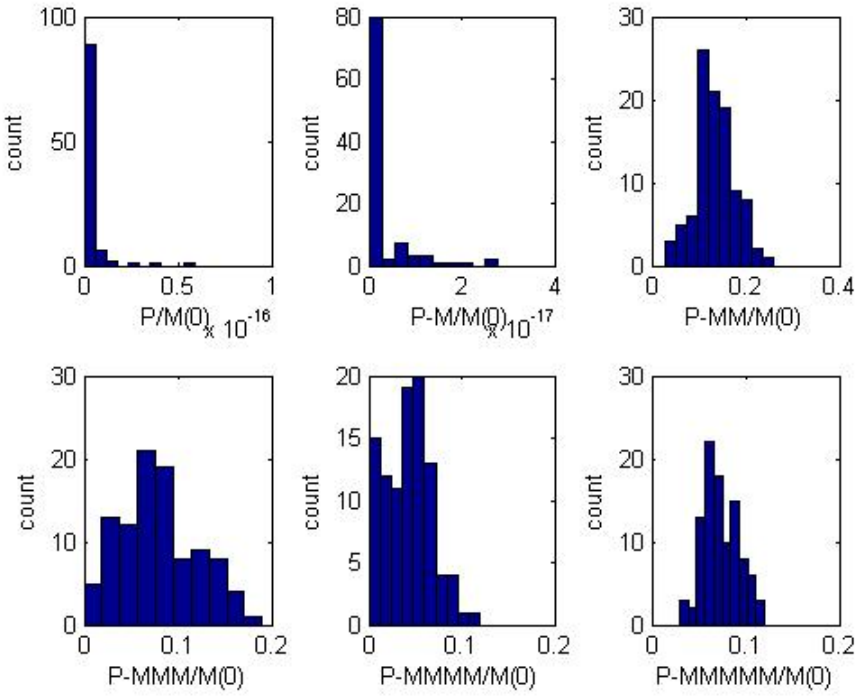

Figure 3.16 – Histograms of the stochastic model results with a ratio of C_P/C_{M_0} of 0,3 and 100 mannose molecules at the entrance of PFR.

Table 3.7 – Composition out of the *Golgi* in a stochastic model with a ratio of C_P/C_{M_0} of 0,3 and 100 mannose molecules at the entrance of PFR.

	Mean	**Standard deviation**	**Percentage of C_{P_0}**
$\mathbf{C_M/C_{M_0}}$	0	0	0
$\mathbf{C_P/C_{M_0}}$	0	0	0
$\mathbf{C_{PM_1}/C_{M_0}}$	0	0	0
$\mathbf{C_{PM_2}/C_{M_0}}$	0,14	0,042	45,20
$\mathbf{C_{PM_3}/C_{M_0}}$	0,078	0,043	25,90
$\mathbf{C_{PM_4}/C_{M_0}}$	0,043	0,025	14,17
$\mathbf{C_{PM_5}/C_{M_0}}$	0,074	0,021	24,57

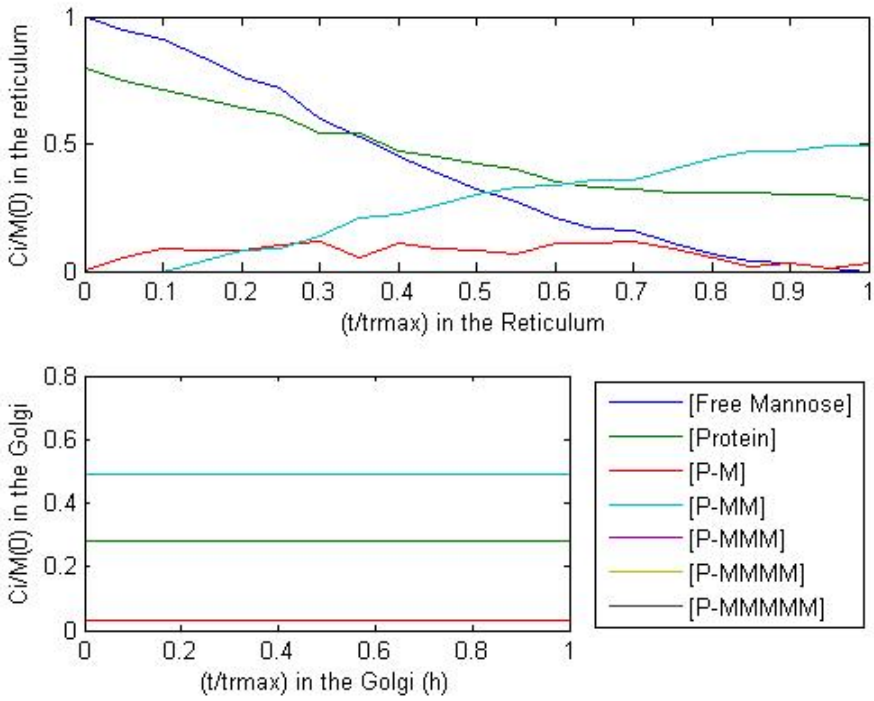

Figure 3.17 – Stochastic model results with a ratio of C_P/C_{M_0} of 0,8 and 100 mannose molecules at the entrance of PFR.

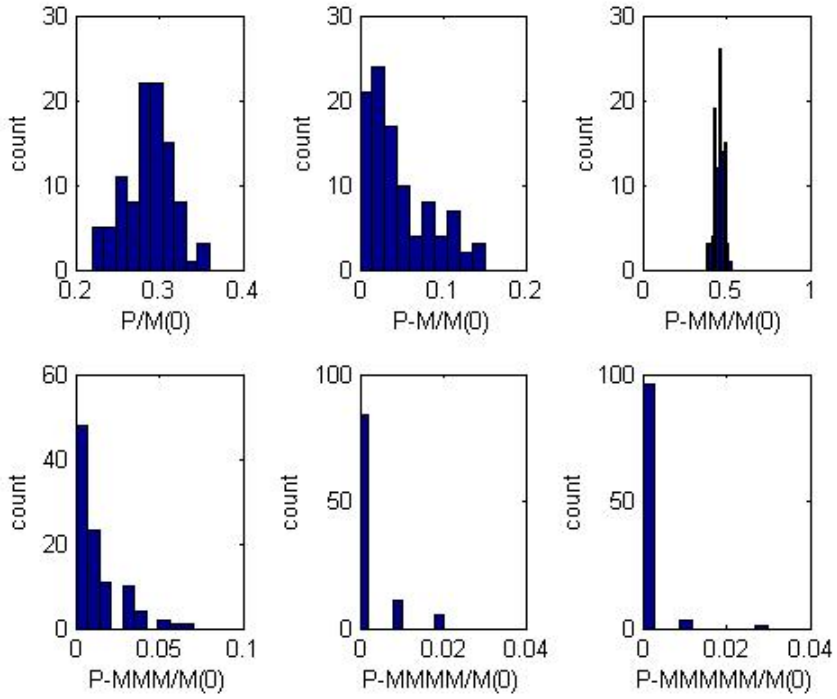

Figure 3.18 – Histograms of the stochastic model results with a ratio of C_P/C_{M_0} of 0,8 and 100 mannose molecules at the entrance of PFR.

Table 3.8 – Composition out of the *Golgi* in a stochastic model with a ratio of C_P/C_{M_0} of 0,8 and 100 mannose molecules at the entrance of PFR.

	Mean	**Standard deviation**	**Percentage of C_{P_0}**
C_M/C_{M_0}	0	0	0
C_P/C_{M_0}	0,29	0,029	36,03
C_{PM_1}/C_{M_0}	0,047	0,039	5,91
C_{PM_2}/C_{M_0}	0,46	0,028	57,53
C_{PM_3}/C_{M_0}	0,011	0,015	1,43
C_{PM_4}/C_{M_0}	0,0021	0,0052	0,26
C_{PM_5}/C_{M_0}	0,0006	0,0034	0,075

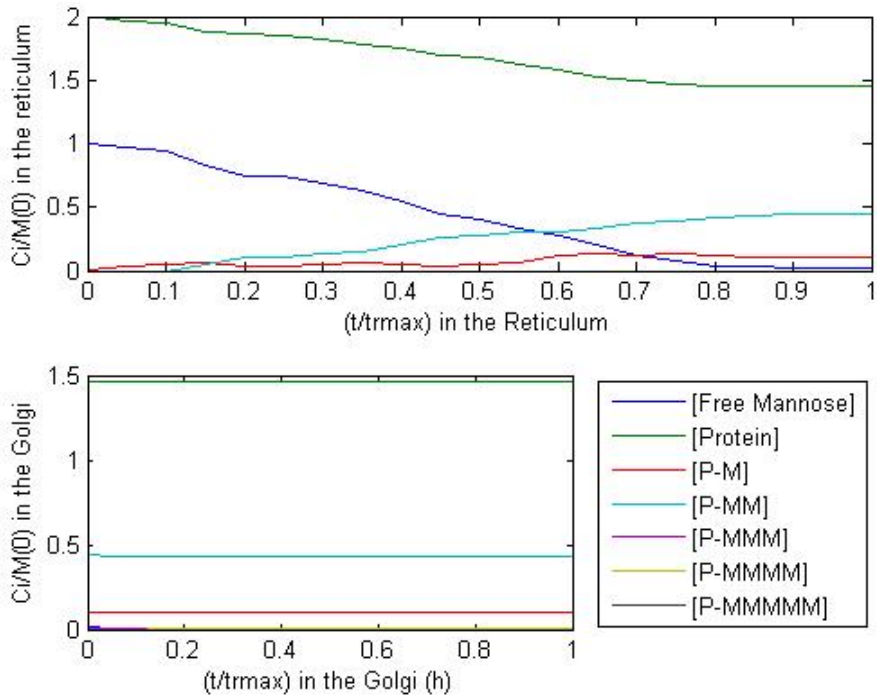

Figure 3.19 – Stochastic model results with a ratio of C_P/C_{M_0} of 2 and 100 mannose molecules at the entrance of PFR.

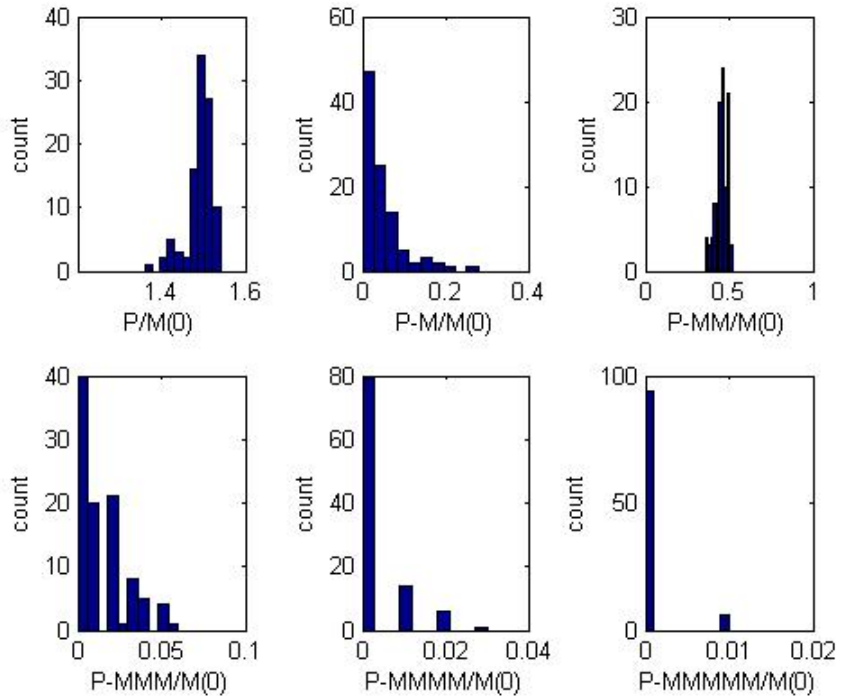

Figure 3.20 – Histograms of the stochastic model results with a ratio of C_P/C_{M_0} of 2 and 100 mannose molecules at the entrance of PFR.

Table 3.9 – Composition out of the *Golgi* in a stochastic model with a ratio of C_P/C_{M_0} of 2 and 100 mannose molecules at the entrance of PFR.

	Mean	**Standard deviation**	**Percentage of C_{P_0}**
C_M/C_{M_0}	0	0	0
C_P/C_{M_0}	1.49	0,023	74,43
C_{PM_1}/C_{M_0}	0,041	0,035	2,07
C_{PM_2}/C_{M_0}	0,47	0,031	23,30
C_{PM_3}/C_{M_0}	0,0092	0,013	0,46
C_{PM_4}/C_{M_0}	0,0025	0,0058	0,13
C_{PM_5}/C_{M_0}	0,0003	0,0017	0,015

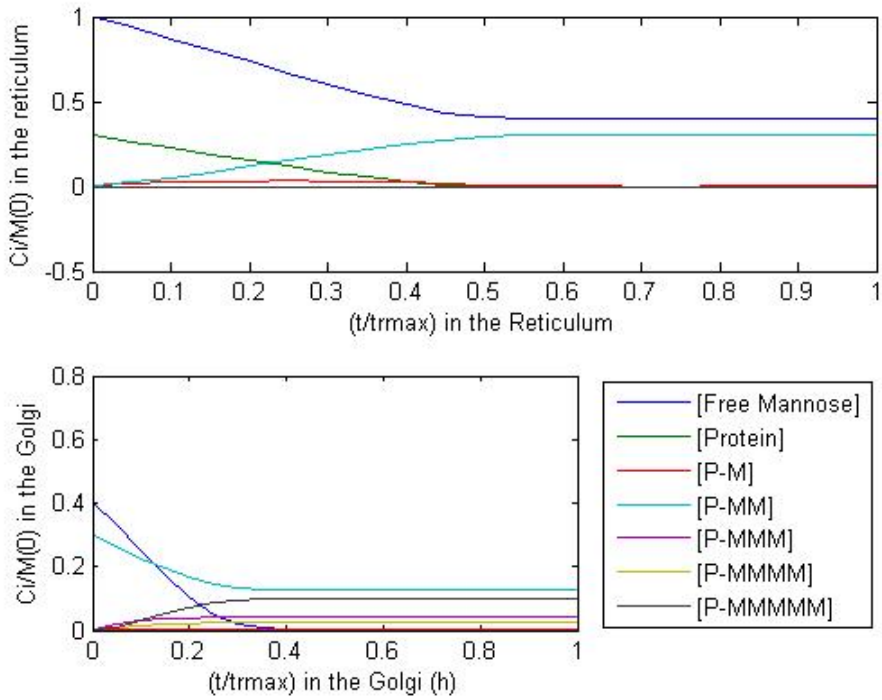

Figure 3.21 – Deterministic model results with a ratio of C_P/C_{M_0} of 0,3; and 5000 mannose molecules at the entrance of PFR.

Table 3.10 – Composition out of the *Golgi* in a deterministic model with a ratio of C_P/C_{M_0} of 0,3 and 5000 mannose molecules at the entrance of PFR.

	Outlow	**Percentage of C_{P_0}**
$\mathbf{C_M/C_{M_0}}$	0	0
$\mathbf{C_P/C_{M_0}}$	0	0
$\mathbf{C_{PM_1}/C_{M_0}}$	0	0
$\mathbf{C_{PM_2}/C_{M_0}}$	0,13	42,65
$\mathbf{C_{PM_3}/C_{M_0}}$	0,045	14,94
$\mathbf{C_{PM_4}/C_{M_0}}$	0,026	8,83
$\mathbf{C_{PM_5}/C_{M_0}}$	0,10	33,58

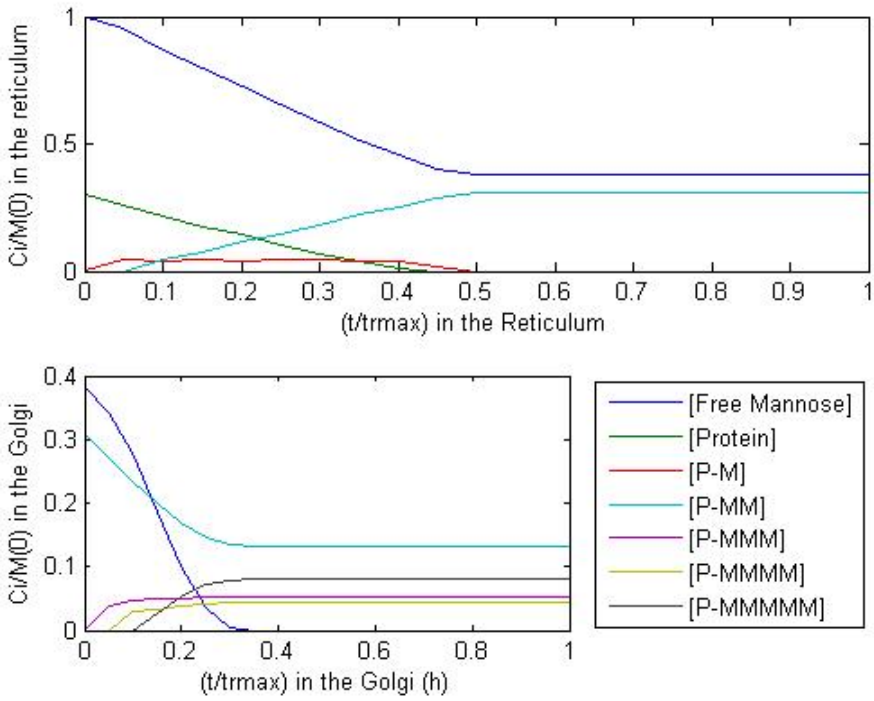

Figure 3.22 – Stochastic model results with a ratio of C_P/C_{M_0} of 0,3 and 5000 mannose molecules at the entrance of PFR.

Table 3.11 – Composition out of the *Golgi* in a stochastic model with a ratio of C_P/C_{M_0} of 0,3 and 5000 mannose molecules at the entrance of PFR.

	Mean	**Standard deviation**	**Percentage of C_{P_0}**
C_M/C_{M_0}	0	0	0
C_P/C_{M_0}	0	0	0
C_{PM_1}/C_{M_0}	0	0	0
C_{PM_2}/C_{M_0}	0.13	0,0076	42,83
C_{PM_3}/C_{M_0}	0,058	0,0064	19,44
C_{PM_4}/C_{M_0}	0,038	0,0054	12,57
C_{PM_5}/C_{M_0}	0,085	0,0037	28,26

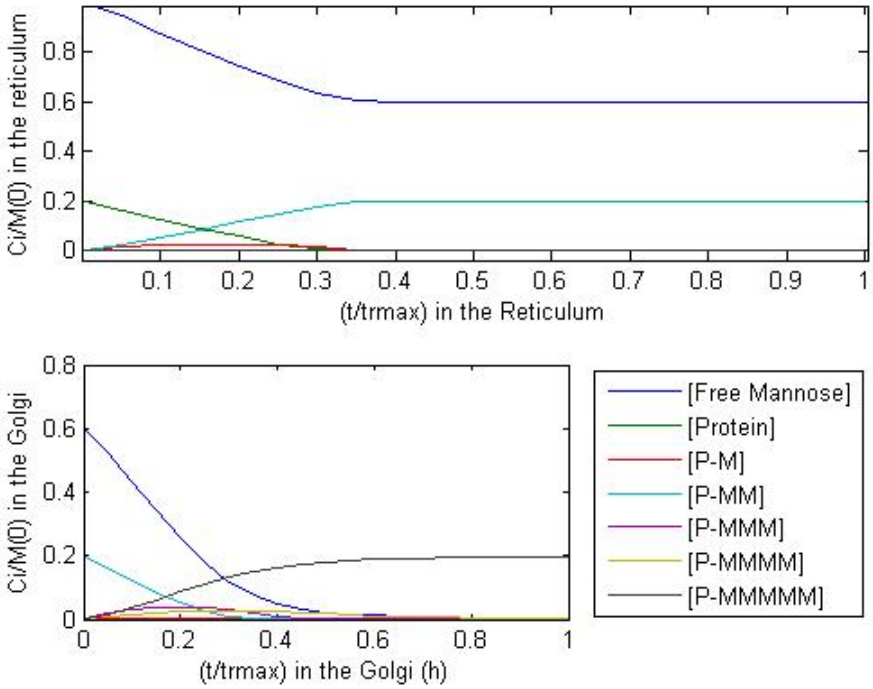

Figure 3.23 – Deterministic model results with a ratio of C_P/C_{M_0} of 0,2 and 30 mannose molecules at the entrance of PFR.

Table 3.12 - Composition out of the *Golgi* in a deterministic model with a ratio of C_P/C_{M_0} of 0,2 and 30 mannose molecules at the entrance of PFR.

	Outlow	**Percentage of C_{P_0}**
$\mathbf{C_M/C_{M_0}}$	0,0031	0.31
$\mathbf{C_P/C_{M_0}}$	0	0
$\mathbf{C_{PM_1}/C_{M_0}}$	0	0
$\mathbf{C_{PM_2}/C_{M_0}}$	0	0,00029
$\mathbf{C_{PM_3}/C_{M_0}}$	0,000077	0,038
$\mathbf{C_{PM_4}/C_{M_0}}$	0,0029	1,46
$\mathbf{C_{PM_5}/C_{M_0}}$	0,20	98,50

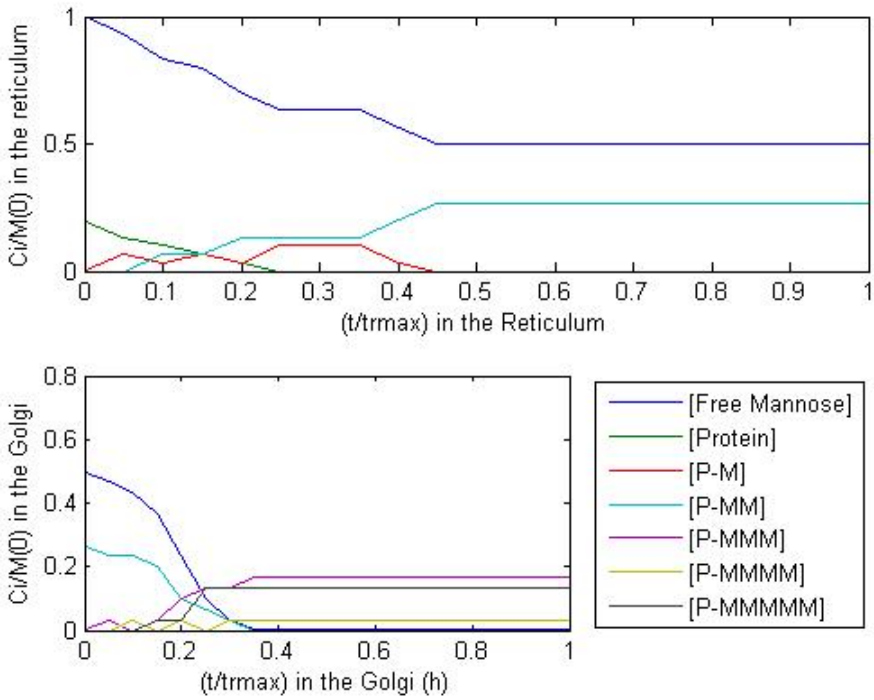

Figure 3.24 — Stochastic model results with a ratio of C_P/C_{M_0} of 0,2 and 30 mannose molecules at the entrance of PFR.

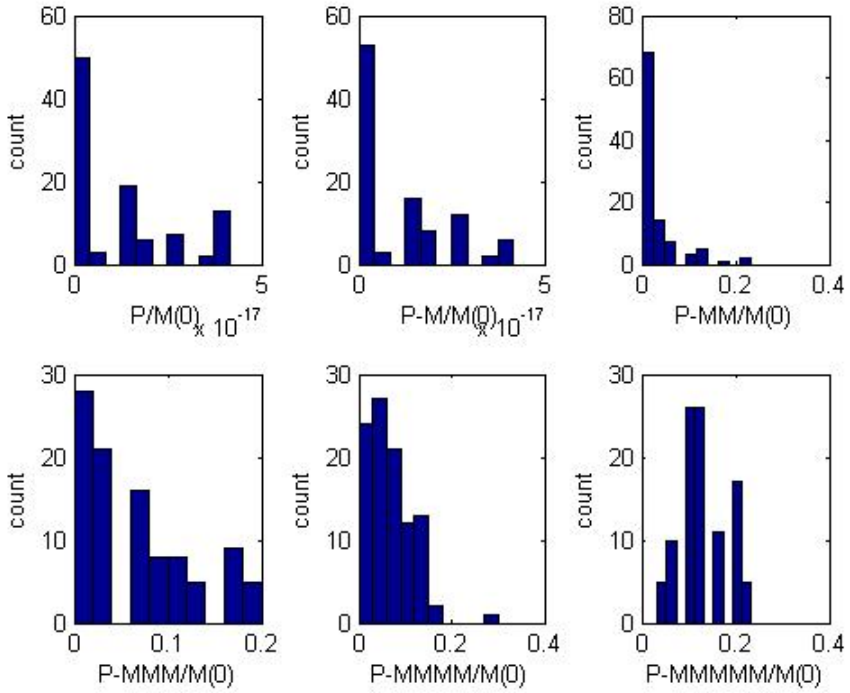

Figure 3.25 – Histograms of the stochastic model results with a ratio of C_P/C_{M_0} of 0,2 and 30 mannose molecules at the entrance of PFR.

Table 3.13 – Composition out of the *Golgi* in a stochastic model with a ratio of C_P/C_{M_0} of 0,2 and 30 mannose molecules at the entrance of PFR.

	Mean	**Standard deviation**	**Percentage of C_{P_0}**
C_M/C_{M_0}	0,00067	0,0047	0,067
C_P/C_{M_0}	0	0	0
C_{M_1}/C_{M_0}	0	0	0
C_{M_2}/C_{M_0}	0,025	0,049	12,67
C_{M_3}/C_{M_0}	0,065	0,06	32,67
C_{M_4}/C_{M_0}	0,058	0,053	29,33
C_{M_5}/C_{M_0}	0,13	0,051	66,50

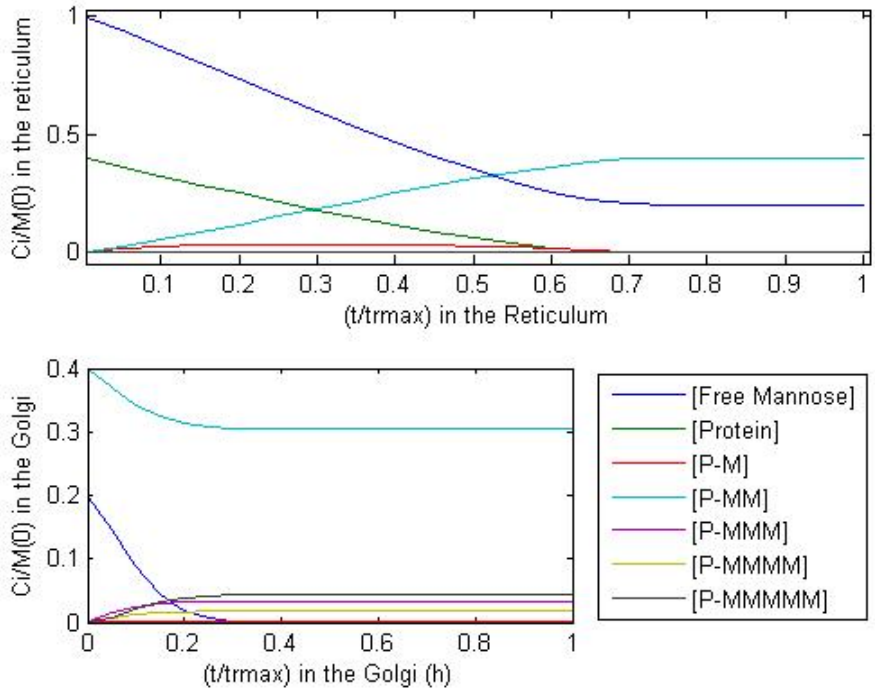

Figure 3.26 - Deterministic model results with a ratio of C_P/C_{M_0} of 0,4 and 30 mannose molecules at the entrance of PFR.

Table 3.14 – Composition out of the *Golgi* in a deterministic model with a ratio of C_P/C_{M_0} of 0,4 and 30 mannose molecules at the entrance of PFR.

	Outlow	**Percentage of C_{P_0}**
C_M/C_{M_0}	0	0
C_P/C_{M_0}	0	0
C_{PM_1}/C_{M_0}	0	0
C_{PM_2}/C_{M_0}	0,31	76,30
C_{PM_3}/C_{M_0}	0,033	8,34
C_{PM_4}/C_{M_0}	0,018	4,42
C_{PM_5}/C_{M_0}	0,044	10,94

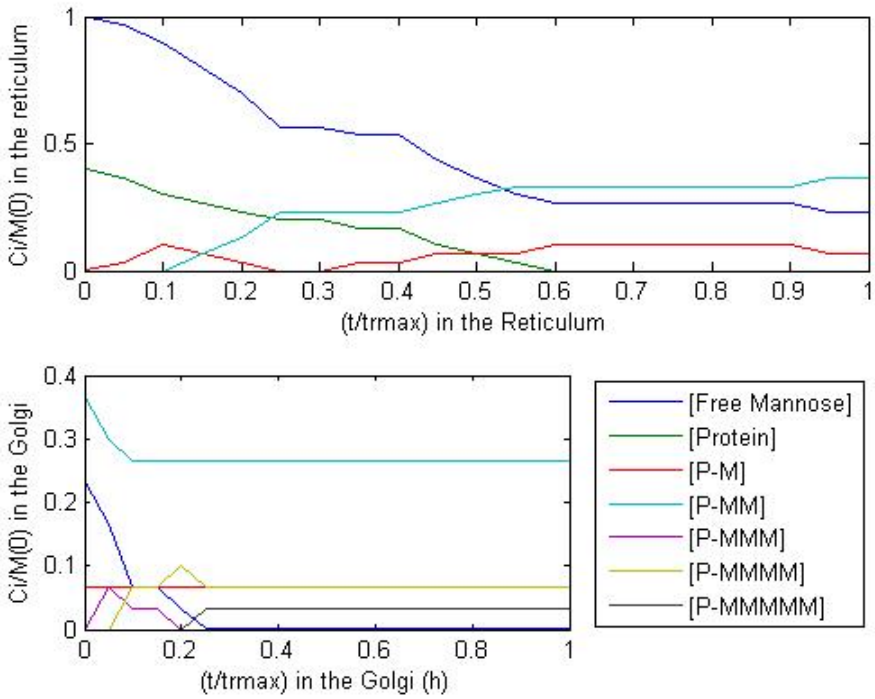

Figure 3.27 – Stochastic model results with a ratio of C_P/C_{M_0} of 0,4 and 30 mannose molecules at the entrance of PFR.

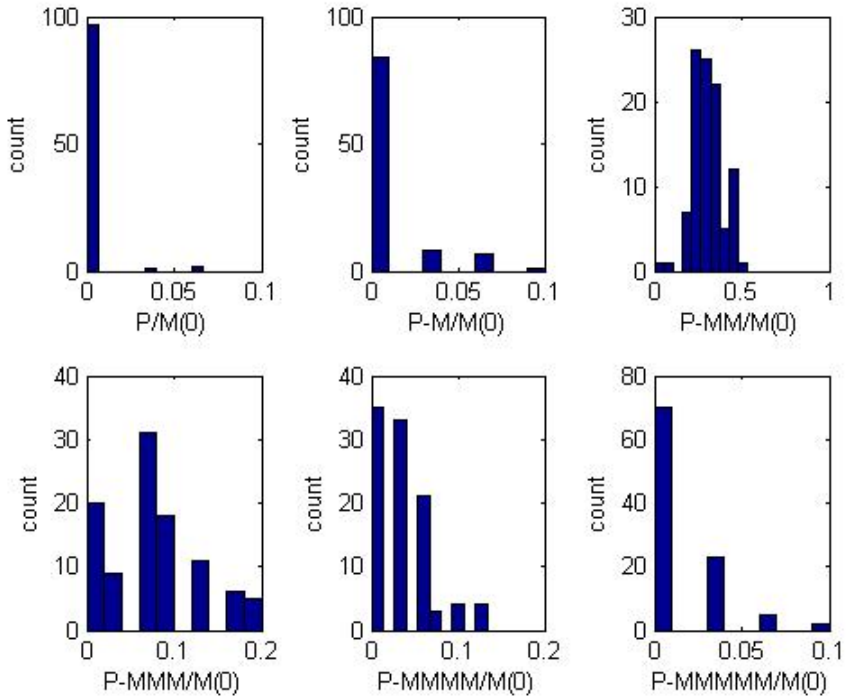

Figure 3.28 – Histograms of the stochastic model results with a ratio of C_P/C_{M_0} of 0,4 and 30 mannose molecules at the entrance of PFR.

Table 3.15 – Composition out of the *Golgi* in a stochastic model with a ratio of C_P/C_{M_0} of 0,4 and 50 mannose molecules at the entrance of PFR.

	Mean	**Standard deviation**	**Percentage of C_{P_0}**
C_M/C_{M_0}	0	0	0
C_P/C_{M_0}	0	0,0099	0,42
C_{PM_1}/C_{M_0}	0,017	0,021	2,08
C_{PM_2}/C_{M_0}	0,31	0,087	76,75
C_{PM_3}/C_{M_0}	0,076	0,056	19,08
C_{PM_4}/C_{M_0}	0,036	0,035	9,08
C_{PM_5}/C_{M_0}	0,013	0,023	3,25

3.1. Glycoprotein Detection Kit

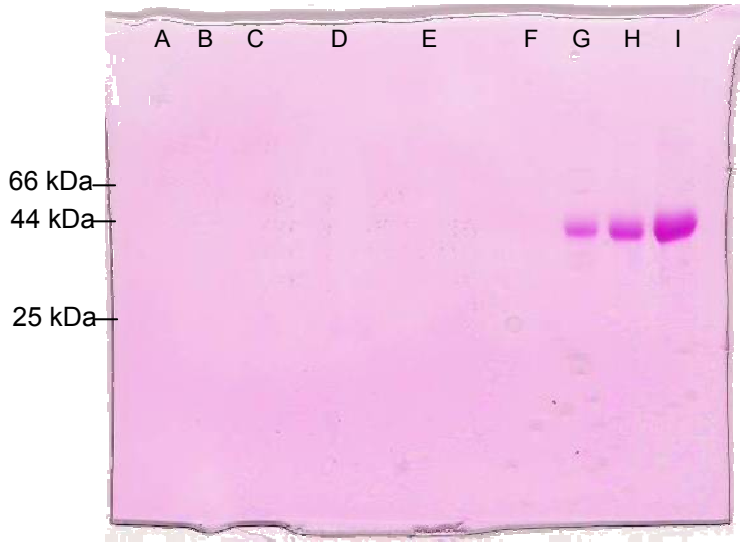

Figure 3.29 – Picture of the Gel with the glycoprotein detection kit. The wells G, H and I are the HPR (Positive standard) with 3µg, 6µg and 12µg respectively.

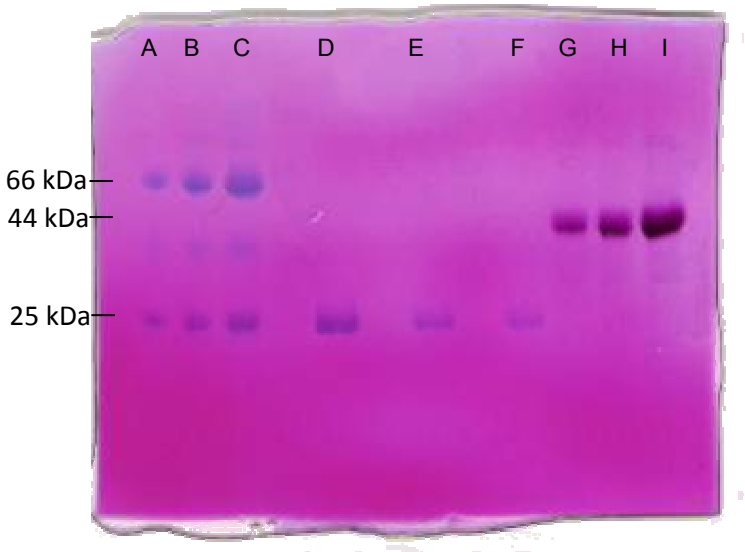

Figure 3.30 - Picture of the Gel with the glycoprotein detection kit and the Simple blue. The wells A, B and C are BSA (Negative control) with 3µg, 6µg and 12µg respectively; D, E and F are scFv with 5,91µg, 11,82µg and 23,64µg respectively; and G, H and I are the HPR (Positive standard) with 3µg, 6µg and 12µg respectively.

4. Discussion

4.1. Fermentation

The first part of this work was the production of the antibody in a pilot scale 50 liters biorector.

The variable airflow rate was set to 30l/min and it remained constant during the fermentation time, showing only small variations, represented in Figure 3.1.

In Figure 3.1, it is represented the temperature, which was constant during the fermentation, however in the phase III it was regulated to the set-point profile defined by Equation 2.1. In the same figure, the pressure is constant during fermentation time.

As stated previously the *P. pastoris* cultivation proceeds in three distinct phases.

In Figure 3.2, the pO_2 decreases during the glycerol batch phase (phase I) until approximately 26 h, with the agitation rate set constant to 300 rpm. During this period the oxygen mass transfer coefficient, k_La is constant, thus the decrease of pO_2 is explained by cell growth that reaches a cell density of approximately 20 g/l at the end of the glycerol batch phase. Once the pO_2 drops to 50%, pO_2 control by the stirrer speed starts automatically resulting in a consistent stirrer speed increase until it reaches the threshold value of 1000 rpm.

To illustrate better the three phases Figure 4.1 shows only the first 60 hours of Figure 3.2. The first phase, as explained above, is the glycerol batch phase. Before the beginning of phase II, it is possible to see the pulse of methanol at approximately 30,67h. Shortly after, the glycerol fed-batch phase starts. In this phase, there is a continuous (exponential) increase of glycerol feeding rate until 43,50 h. The pO_2 is controlled to 50% by manipulating the stirrer speed.

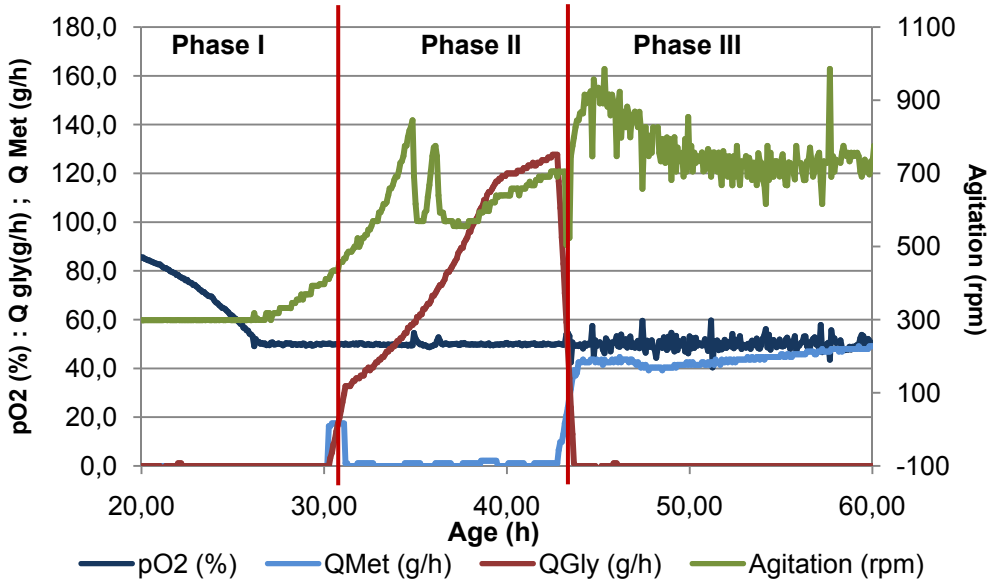

Figure 4.1 – Zoom of the Figure 3.2. Carbon sources feeding, glycerol and methanol; pO_2 and agitation between 20h-60h of fermentation. In this figure is also represented the three distant phases of the fermentation.

Phase III is the methanol fed-batch phase beginning at 43 h,. The methanol feeding rate starts with the value of 40g/h and it increases continuously until the end of the cultivation, reaching then the value of ≈200g/h (this can be observed in Figure 3.2). The pO_2 between 43h to 73h is kept between 40%-60% by stirrer speed manipulation. After 73 h, the pO2 decreases to 32% and the agitation reaches the maximum value of ≈1000rpm.

The Figure 3.3 shows that phase I and phase II has pH 5 while phase III was run at pH ≈ 6,8. The product builds up, with significant levels, only in phase III – the methanol fed-batch phase, because the methanol induces the production of the product. Until 70h, the product has a continuous increase but after that it suffers a decrease and remains between 3-4mg/l for 75-105h. Afterwards, the product shows another continuous increase until the end of the cultivation.

The behavior of the product was not similar to the behavior of the biomass growth (see Figure 3.4). The behavior of the cellular growth was analyzed by OD and the wet cell weight. These two parameters show that the cells have a higher growth rate in the first 50 h (phases I and II) whereas in phase III cell growth is maintained but at a lower rate. There is no visible decrease in the biomass concentration as opposed to the product concentration; thus the product is most likely non growth associated. The decrease in product mass is probably caused by proteolytic cleavages due to proteases excreted to the medium (Jahic, *et al*., 2003).

4.2. Proteins Assay Kit

After purification, the product was analyzed by MALDI TOF/TOF MS. The product had to fulfill several requirements, prior to MALDI analysis, in order to ensure reliability of the final results (Lennon, 1997).

A potential problem for MALDI analysis was the high concentration of glycerol in the product solution. After purification, the product was stored in PBS with 10% of glycerol at 7,4 pH. The glycerol is added in the purification because this compound works as a stabilizer (it maintains the product intact for longer periods of time without degradation).

The first step was to make the concentration of glycerol acceptable for MALDI analysis. It was chosen to do the substitution of the glycerol by a buffer which was tolerated by MALDI (the PBS). The Amicon Ultra-4Centrifugal Filter was chosen because it is a simple and cheap method and because it enables concentration sample at the same time.

After concentration and glycerol substitution, it was performed the BCA protein Assay Kit. With this assay, it is always necessary to do a calibration curve with BSA (the standard protein). The concentrations were determined only with the values between the parameters of the calibration curve – these values were the ones related to the dilution factor of 8 to 64. The respective standard deviations were calculated. The following table shows the concentration average of the product after purification and after using the Amicon Ultra-4Centrifugal Filter (Table 4.1). The solution without glycerol was deliberately more concentrated because the next stage was the MALDI analysis.

Table 4.1 – Average concentrations of the product after purification and after using the Amicon Ultra-4Centrifugal Filter and their standard deviations.

Product after purification	Product after concentration and substitution of glycerol
2,955 ± 0,141 mg/ml	4,029 ± 0,363 mg/ml

4.3. MALDI and Bioinformatic methods

Just before MALDI analysis the only information available about the target protein was: i) the protein is called AP039, ii) it is an antibody fragment of Anti-(ED-B) scFv, and the protein could have two different types of structures.

In the beginning, it was decided to do three MALDI analyses: the first one was done to the intact protein; the second one was done over the digested product by a specific enzyme; and the third was with the same sample of the second but with one additional enzyme – a specific glycoprotein enzyme.

The first MALDI analysis was done with the intact protein and the result is shown in the Figure 3.6. In this MALDI spectrum it can be observed three significant peaks. However, when the peaks are observed indetail – Annex C – it can be seen that the peaks are not smooth, presenting some irregularities. These irregularities can mean either that apart from the three different structures, it may exist other very similar structures, or that they are the same structures but the MALDI analysis induced some kind of degradation.

The second spectrum (Figure 3.7) corresponds to the protein digested by trypsin, which cuts the amino acids sequence at specific places – it cuts at lysine or arginine except if they are followed by proline. The MALDI data were run through NCBI (National Center for Biotechnology Information) database to find out the protein sequence. The results from the NCBI database reported that the protein with highest score was the Anti-(ED-B) scFv (*Homo sapiens*) with the accession number of gi|3395510. The data of MALDI were also supplied[5] in Table 7.5.

The third analysis was made after the additional digestion by the PNGase F – this enzyme cuts the N-linked glycans from the peptide where they were attached. So, it was expected, with these two analyses, to detect the peptide sequence which contains the glycans and the glycans itself. The spectrum, Figure 3.8, has more peaks in comparison with Figure 3.7. The data of MALDI were also supplied[5] in Table 7.6.

After this direct review of the data provided by the Mass Spectrometry Laboratory, it was made an intense study.

One of the goals with these analyses was to determine the aminoacid sequence of AP039. It was used also the Database of NCBI (NCBI) in the format of GenBank.

[5] Data provided/obtained by the Mass Spectrometry Laboratory, Analytical Services Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa.

The Database provided the L19 anti-(ED-B) scFv (Homo sapiens) aminoacid sequence after translation, with the accession number gi|3395509.

The name L19 is different than AP039, but it was established in the beginning that the protein produced was the latter. So it was made a literature research to clarify this issue.

The AP039 has an aminoacid sequence of the L19 anti-ED-B scFV plus five aminoacids in C-terminal. The AP39 anti-ED-B scFV is also used for radiolabeling for molecular imaging of the solid tumor. This sequence of five aminoacids produces a more stable radiolabeling binding (Berndorff, *et al*., 2006). The sequence of AP039 that is described in Figure 3.9 is the same as the sequence of L19 – Figure 7.7 (Annex E) – plus five more aminoacids that are added in the C-terminal.

Another important piece of information about this protein has to do with its structure: the L19 can have the structure of an associative dimer, where the scFv interaction is only by Wan der Vaals forces; or it can have the structure of covalent dimer, where the V_L's are attached with disulfide bonds – Figure 7.6 (Annex E) (Berndorff, *et al*., 2005). It is likely that the AP039 can same the structural forms. This hypothesis is consistent with the MALDI analysis results of Figure 3.6: the peak with 25736 m/z can be AP039 in monomer form; the 51186 m/z peak can be AP039 in dimer form - covalent or associative; and the 12914 m/z peak can be only the V_H or V_L of the AP039.

It is also stated in the literature that the homodimer forms are efficiently radiolabeled whereas the monomeric form is not. It is possible to convert the homodimeric form in the monomeric form (under reducing conditions), and interestingly the resulting monomers are also efficiently radiolabeled. So, this information suggests that in the antibody production process, there may be different events in the formation of the homodimers and the monomers, like different folding, and possibly glycosylation, among other (Cunha, *et al*., 2004).

The sequence of AP039 was run through the PeptideCutter and PeptideMass bioinformatic tools (see bioinformatic methods). These tools predicted the peptide sequence of trypsin cleavages and gave their respective mass/charge (m/z) – Table 3.2. The data from MALDI analysis – Table 7.7 (Annex F) – is used to compare with Table 3.3, which allows identifying only six peaks – Table 4.2. The minimum value of m/z is 775 so the peptide sequence with lower m/z does not appear in the MALDI

spectrum. The maximum value is 4012 m/z so the peak of 6177,8821 m/z does not appear. Therefore, the proportion six out of nine is a very reasonable value.

Table 4.2 - List of the six peaks identified by comparing the results shown in Table 7.5/Table 7.6 to Table 3.3.

Theoretical peak	Experimental peak	Δmass (Dalton)	Position	Sequence
1085,5989	1085,6067	0,0078	177-185	LLIYYASSR
1352,6991	1352,6945	-0,0046	77-87	NTLYLQMNSLR
1632,7864	1632,7731	-0,0133	193-208	FSGSGSGTDFTLTISR
1896,0185	1895,9832	-0,0353	1-19	EVQLLESGGGLVQPGGSLR
1918,8640	1918,8375	-0,0265	209-224	LEPEDFAVYYCQQTGR
2423,2102	2423,1599	-0,0503	155-176	ASQSVSSSFLAWYQQKPGQAPR

As explained in the introduction, the sequence for N-linked glycosylation is Asn-X-Ser/Thr (N-Xaa-T/S) and for O-linked glycosylation is only Ser/Thr (S/T). In Figure 7.8 (Annex E) it can be seen that there is no sequence of N-linked glycosylation and the O-linked glycosylation have fifty-six possible spots for attachment.

Applying the GlycoMod tool was the next step. This tool can predict the N-linked or O-linked glycoforms from the data obtained from MALDI along with the known aminoacid sequence.

In this case and given the information above, the O-linked glycans were the only ones analyzed. It was used the data from MALDI with protein cleaved by trypsin and cleaved by trypsin and PNGase F. Table 7.7 (Annex F) lists all possible O-linked glycans. It is made one first selection – remove all structures with more than five mannoses – Table 4.2 – since the *Pichia pastoris* can only have O-linked glycans with maximum of five linear mannoses (see 1.4.2 Processing of O-linked Oligosaccharide).

Table 4.2 shows the Data of both cleavage enzymes (Table 7.5 and Table 7.6), so it is important to look over the results one by one. The results with position 77-87, 209-225 with the structure $(Man)_4$, both 66-76 and 225-238 only appear in the third spectrum (Figure 3.8); so it is likely that these values correspond to degraded sequences. The sequence with position 177-185 with the structure $(Man)_5$ and 225-243 appear in both spectra (Figure 3.7 and Figure 3.8). However, these masses are identified as aminoacid sequences (Table 7.5).

The remaining possible glycoforms are presented in Table 4.3. The sequence with the position 225-235 only appears in Figure 3.7, it is possible that this sequence had suffered degradation over time and didn't appear in Figure 3.8. The other sequences appear in both spectra (Figure 3.7 and Figure 3.8) so it is possible to have this structures in the sample.

Table 4.3– Possible glycoforms by GlycoMod.

Position	Sequence	Experimental mass (m/z)	Glycoform mass (m/z)	Δmass (Dalton)	Peptide mass	Structure
177-185	LLIYYASSR	1409,7137	324,106	0,009	1084,592	(Man)2
209-224	LEPEDFAVYYCQQTGR	2079,9773	162,053	-0,939	1917,857	(Man)1
225-235	IPPTFGQGTK	1368,6901	324,106	-0,983	1044,56	(Man)2

4.4. Glycosylation mathematical models

Simulations of the deterministic and stochastic mathematical models were performed in order to assess the likely glycosylation dynamics and final glycoforms distribution. The plan was to compare these theoretical distributions with the experimental distributions provided by MALDI.

It is very important to refer that both programs were based in glycosyltransferase kinetic parameters that are unknown and unfortunately it was not possible to obtain these values in public literature sources. So, the maximum rate constants, expressed in dimensionless form (units of fraction of linked mannose per unit of residence time in the *Golgi* or E.R.) can be viewed as the relative strength of enzymes. The model is thus useful to study the sensitivity of glycoforms distribution to the strength of individual glycotransferases. The model can obviously be tuned to the posterior MALDI data.

To compare the results of the deterministic model it was analyzed different ratios between protein concentration and initial concentration of mannose (C_P/C_{M_0}). When the ratio is 0,3 (Figure 3.10) in E.R. the species of protein with one and two mannose (PM_1 and PM_2) were formed while free mannose decreases and free protein

depletes completely. That is, all proteins were glycosylated and moreover, the PM_1 was completly transformed into PM_2. Furthermore, in the *Golgi*, the species of PM_2 was completely consumed and all of the free mannoses were depleted resulting in the other species (PM_3, PM_4 and PM_5) were formed until exhaustion of free mannoses. The results that appear in the table in the deterministic model are directly associated with the figure, since this model produce the same output for a given starting condition. The results of Table 3.4 suggested that in the end of the glycosylation the predominant species is PM_2, followed by PM_5.

When the ratio C_P/C_{M_0} is 0,8 (Figure 3.11), in E.R., the PM_1 and PM_2 were formed, the protein was consumed and almost all the free mannose was depleted. In the *Golgi* (Figure 3.12), the species of PM_2 decreased and the other species were formed until the exhaustion of all free mannoses. In the end of glycosylation (Table 3.5) the prevalent species was PM_2 followed by free protein.

The result with a ratio C_P/C_{M_0} of 2 was very similar to the previous ratio, with the exception that a great deal of proteins are not glycosylated. In the end (Table 3.6) the results were the opposite of the previous ratio – the prevalent species was the free protein, followed by PM_2.

To compare the results of the stochastic model it was analyzed different ratios between protein concentration and initial concentration of mannose (C_P/C_{M_0}). It is important to realize that all the results of the stochastic model had minor fluctuations in the general tendency of the graphics. Like it was referred in the methods, each time the program of stochastic model is executed it provides different results, since this method is associated with randomness.

In the E.R. when the ratio C_P/C_{M_0} was 0,3 (Figure 3.15), the PM_1 and PM_2 were formed, there is a decrease in free mannose and all protein was consumed. In the *Golgi*, the species PM_2 decreased and the other species were formed until the exhaustion of free mannoses. The results that appear in table in the stochastic model are associated with the figure of the histograms, where is possible to observe the randomness of the model. The Figure 3.16 and Table 3.7 demonstrate that the predominant species were the PM_2, followed by PM_3 and PM_5.

Figure 3.17 (C_P/C_{M_0} of 0,8) presents, in the E.R., the formation of PM_1 and PM_2, the consumption of protein and the depletion of free mannose, given that in the *Golgi* all species remain constant and there were not any new species formation. In Figure

3.18 and Table 3.8 the species PM_2 was predominant, followed by free protein. These results also demonstrated that the formation of PM_3, PM_4 and PM_5 was negligible.

Likewise, the result with the ratio C_P/C_{M_0} of 2 (Figure 3.20. and Table 3.9) were very similar to the corresponding deterministic model results.

Two groups of simulations were performed to compare the deterministic and stochastic models: one group with $N_{MAN,0}$=5000 mannose molecules and another with $N_{MAN,0}$=30 mannose molecules at the entrance of the PFR. In the former case, for a ratio P/M_0 of 0,3, the predictions of both models (Figure 3.21 and Figure 3.22; Table 3.10 and Table 3.11) were very similar. The predominant species, in the outflow of the *Golgi*, was PM2, followed by PM_5. The fluctuations that appeared in all of the previous figures of the stochastic model disappeared and the behavior of stochastic model was very similar to the deterministic model. It is possible to conclude that for high number of mannose molecules at the entrance of PFR ($N_{MAN,0}$) the stochastic model results converge to the ones of the deterministic model.

For the latter case, when the number of mannose molecules is 30, the deterministic model predicts that for a ratio C_P/C_{M_0} of 0,4 the predominant species is PM_2, followed by PM_5 (Figure 3.26 and Table 3.14). However, the stochastic model predicts for the same C_P/C_{M_0} ratio that the predominant species is PM_2, followed by PM_3 (Figure 3.27 and Table 3.15). Decreasing further the ratio C_P/C_{M_0} ratio to values lower than 0,2, the deterministic model with 30 mannoses molecules at the entrance predicts that PM_5 is the predominant species, followed by PM_4 (Figure 3.23 and Table 3.12). Again, the stochastic model predicts different predominant species, i.e. first PM_5 followed by PM_3 (Figure 3.24 and Table 3.13).

4.5. Glycoprotein Detection Kit

The next step was assuring that these structures really exist by MALDI analysis and to quantify the relative mass fraction of the several glycoforms and to compare with the predictions of the mathematical models. The preparation for MALDI involves releasing the O-linked glycans, which is a complicated method and has a great probability of degrading the glycoforms in the processing steps (example in Annex G). Therefore, it was decided to do first the glycoprotein detection kit.

This kit is a simple and easy method to detect glycoproteins. First it was made a PAGE, where the spots A, B and C had the BSA protein – the negative control – because BSA is not glycosylated. The spots G, H and I had the HPR protein which has

16% of glycans – so it was the positive control. The spots D, E and F had a sample of the protein produced.

The Figure 3.29 has the Glycoprotein Detection Kit treatment and only appears the positive control. Since the other proteins did not appear it could generate some doubt: the protein didn't appear in the gel or the protein wasn't a glycoprotein. So the gel was colored with Simple Blue, this component dyes all proteins – Figure 3.30. It can be seen proteins in all spots, so it is possible to conclude that the protein produced wasn't a glycoprotein.

5. Conclusion

The first part of this study was the production of Anti-(ED-B) scFv by *Pichia pastoris* in a 50 L fermenter. The production of Anti-(ED-B) scFv takes place during phase III of the fermentation, with the maximum concentration obtained at the end of phase III being 8,4mg/L. This concentration is very small, so in the future it is necessary to do more optimizations of phase III operation, namely of the methanol feeding strategy, temperature and pH and eventually media supplementation with key nutrients.

The MALDI analysis was very important to identify accurately the protein that was produced. The information of three different structures was proved by MALDI, and the information of conversion of the homodimeric form to the monomeric form implies that there is posttranslational process in the production of this antibody.

Just from the peptide sequence it was possible to predict several structures using protein bioinformatics. When the MALDI and bioinformatic data of the antibody were analyzed in detail, it was found that there is no sequence for N-linked glycosylation, which was expected, and that for the sequence of O-linked glycosylation there are 56 possible spots. However this peptide sequence can only possess three structures, being one of them repeated (the possible structures are Man_2, Man_1).

Two mathematical models were implemented in MATLAB to study glycosylation dynamics and to predict the distribution of glycoforms and eventually to compare with the MALDI data. There is no data about the in vivo activity of the involved glycosyltransferases thus the scope of these models is restricted to a sensitivity analysys of the final glycoform distribution over the relative activity of enzymes, the ratio of protein to mannose, and total amount of mannose available for glycosylation, which is supposed to be the limiting factor. It was concluded that when the Protein/Mannose ratio is low the prevalent glycoforms are PM_2. While when the protein/mannose ratio is very low, the prevalent glycoforms are PM_5. These results are rational but obviously the enzyme activities play an important role on the exact distribution. It was also verified that for high number of mannose molecules at the beginning of the glycosylation process there is a convergence of results predicted by the deterministic and stochastic models. However, both models diverge when the number of mannose molecules is very low. In this situation, the stochastic predictions

are more consistent with the true nature of the system than the deterministic ones. It can be generally observed a higher dispersion in the glycoforms distribution. Higher structures, such as PM_5, have a higher weight even for relatively high protein/mannose ratios.

Finally, it was important to validate the distribution of glycoforms experimentally. However, the Glycoprotein Detection Kit showed that the anti-(ED-B) scFv produced by *Pichia pastoris* at the culture conditions performed, using methanol as carbon source, is not glycosylated. Apparently, the posttranslational modifications that lead to the formation of monomers or dimers are not related to glycosylation but probably to the folding process, in these experimental conditions.

In future developments, it is necessary to consider the following issues:

- To change experimental conditions (pH, Temperature, methanol feeding rate) or strain (for instance to a constitutive strain grown on glycerol) or, if necessary, the expression vector to a protein with guaranteed N- or O-glycosylation;
- To evaluate the type of glycans that are attached in the protein by MALDI analysis, with the help of protein Bioinformatics;
- To find the best technique for estimating the glycans over the fermentation time – in this respect there are many possible analytical techniques (HPAEC-PAD, HPLC, GC…);
- To develop and validate the mathematical model by comparing model predictions and MALDI glycoform measurements.
- To assess glycosylation controllability through manipulation of environmental conditions or by genetic engineering targeting the increase mannose turnover, the most likely bottleneck, or higher expression levels of glycotransferases.

6. References

AguardLab. Aguard Lab. [Online] http://www.msg.ucsf.edu/agard/protocols.html.

Bailon, Pascal, *et al*. 2000. *Affinity Chromatography - Methods and Protocols.* s.l. : Humana Press, 2000.

Berndorff, Dietmar, *et al*. 2006. Imaging of Tumor Angiogenesis Using 99mTc-Labeled Human Recombinat Anti-ED-B Fibronectin Antibody Fragments. *The Journal of Nuclear Medecine.* 2006, Vol. 37, 10, pp. 1707-1716.

Berndorff, Dietmar, *et al*. 2005. Radioimmunotherapy of Solid Tumors by Targeting Extra Domain B Fibronectin: Identification of the Best-Suited Radioimmunoconjugate. *Clinical Cancer Research.* 2005, Vol. 11, 19, pp. 7053-7063.

Bretthauer, Roger K. and Castellino, Francis J. 1999. Glycosylation of Pichia pastoris - derived proteins. *Biotechnology and Applied Biochemistry.* 1999, Vol. 30, pp. 193-200.

Burda, Patricie and Aebi, Markus. 1999. The dolichol pathway of N-linked glycosylation. *Biochimica et Biophysica Acta 1426.* 1999, pp. 239-257.

Corsi, Ann K. and Schekman, Randy. 1996. Mechanism of Polypeptide Translocation into the Endoplasmatic Reticulum. *The Journal of Biologoical Chemistry.* 1996, Vols. 271, Nº 48, 29, pp. 30299-30302.

Cos, Oriol, *et al*. 2006. A Simple Model-Based Control for Pichia pastoris Allows a More Efficient Heterologous Protein Production Bioprocess. *Biotechnology and Bioengineering.* 2006, Vol. 95, 1, pp. 145-154.

Crowther, John R. 2009. *The ELISA Guidebook.* Second Edition. Austria : Humana Press, 2009.

Cunha, A. E., *et al*. 2004. Methanol Induction Optimization for scFv Antibody Fragment Production in Pichia pastoris. *Biotechnology and Bioengineering.* 2004, Vol. 86, 4.

d'Anjou, M. C. and Daugulis, A. J. 1997. A model-based feeding strategy for fed-batch fermentation of recombinant Pichia pastoris. *Biotechnology Techniques.* 1997, Vol. 11, 12, pp. 865–868.

d'Anjou, Marc C. and Daugulis, Andrew J. 2000. Mixed-feed exponential feeding for fed-batch culture of recombinant methylotrophic yeast. *Biotechnology Letters.* 2000, Vol. 22, pp. 341–346.

Damasceno, Leonardo M., *et al*. 2004. An optimized fermentation process for high-level production of a single-chain Fv antibody fragment in Pichia pastoris. *Protein Expression and Purification.* 2004, Vol. 37, pp. 18–26.

Deepak, A. V., *et al*. 2003 . Isolation and characterization of a 29-kDa glycoprotein with antifungal activity from bulbs of Urginea indica. *Biochemical and Biophysical Research Communications.* 311, 2003 , pp. 735–742.

ExPASy, The. http://www.expasy.org/tools/ . [Online]

Gasteiger, Elisabeth, *et al*. 2005. Protein Identification and Analysis Tools. [book auth.] John M. Walker. *The Proteomics Protocols Handbook.* s.l. : Humana Press, 2005, pp. 571-607.

Gillespie, Daniel T. 2005. Stochastic Chemical Kinetics. [book auth.] S Yip. *Handbook of Materials Modeling.* s.l. : Spinger , 2005, pp. 1735-1752.

Glick, Bernard R. and Paternak, Jack J. 2003. *Molecular Biotechnology - Principles and application of recombinant DNA.* 3ª Edition. USA : ASM Press Washington, DC, 2003. pp. 163-174.

Gonzalez, R., Andrews, B. A. and Asenjo, J. A. 2001. Metabolic Control Analysis of Monoclonal Antibody Synthesis. *Biotechnology Progress.* 2001, Vol. 17 , 2, pp. 217-226.

Goto, Masctoshi. 2007. Protein O-Glycosylation in fungi: diverse structures and multiple functions. *Bioscience, Biotechnology and Biochemistry.* 2007, Vol. 71, 6, pp. 1415-1427.

Hamilton, Stephen R. and Gerngross, Tillman U. 2007. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Current Opinion in Biotechnology .* 2007, Vol. 18, pp. 387-392.

Helenius, Ari and Aebi, Markus. 2004. Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review Biochemistry.* 2004, Vol. 73, pp. 1019-1049.

Hellwig, Stephan, *et al*. 2001. Analysis of Single-Chain Antibody Production in Pichia pastoris Using On-Line Methanol Control in Fed-Batch and Mixed-Feed Fermentations. *Biotechnology and Bioengineering.* 2001, Vol. 74, 4, pp. 344-352.

Herscovics, Annette and Orlean, Peter. 1993. Glycoprotein biosynthesis in yeast. *FASEB Journal .* 1993, Vol. 7, 6, pp. 540-550.

Herscovics, Annette. 1999. Processing glycosidases of Saccharomyces cerevisiae. *Biochimica et Biophysica Acta 1426.* 1999, pp. 275-285.

Hillenkamp, Franz and Peter-Katalinic, Jasna. 2007. *MALDI MS - A pratical guide to instrumentation methods and applications.* Germany : Wiley-VCH, 2007.

Hossler, P., *et al*. 2006. GlycoVis: visualizing glycan distribution in the protein N-glycosylation pathway in mammalian cells. *Biotechnology and bioengineering .* 2006, Vol. 95, 5, pp. 946-960.

Invitrogen®. http://tools.invitrogen.com/content/sfs/manuals/ppic9k_man.pdf. [Online] [Cited: March 17, 2009.]

Jahic, M., *et al*. 2002. Modeling of growth and energy metabolism of Pichia pastoris producing a fusion protein. *Bioprocess Biosyst. Eng.* 24 , 2002, pp. 385-393.

Jahic, Mehmedalija, *et al*. 2003. Temperature limited fed-batch technique for control of proteolysis in Pichia pastoris bioreactor cultures. *Microbial Cell Factories.* 2003, Vol. 2.

Kukuruzinska, M. A. and Lennon, K. 1998. Protein N-Glycosylation: Molecular Genetics and Functional Significance. *Critical Reviews in Oral Biology & Medicine.* 1998, Vol. 9, 4, pp. 415-448.

Kukuruzinska, M. A., Bergh, M. L. E. and Jackson, B. J. 1987. Protein glycosylation in yeast. *Annual Reviews in Biochemistry.* 1987, Vol. 56, pp. 915-944.

Lee, Marcus C.S., *et al*. 2004. Bi-Directional Protein Transport between the ER and Golgi. *Annual Review of Cell and Developmental Biology.* 2004, Vol. 20, pp. 87-123.

Leibiger, Haike, *et al*. 1999 . Variable domain-linked oligosaccharides of a human monoclonal IgG: structure and influence on antigen binding. *Biochem. J. .* 338 , 1999 , pp. 529–538.

Lennon, John J. 1997. Matrix Assisted Laser Desorption Ionization Time-of-flight Mass Spectrometry. [Online] 1997. [Cited: Jan 13, 2009.] http://www.abrf.org/ABRFNews/1997/June1997/jun97lennon.html.

Lodish, Harvey, *et al*. 1999. *Molecular Cell Biology.* Fourth Edition. USA : Media Connected - W. H. Freeman and Company, 1999.

Milipore, Immunodectection -. http://www.millipore.com/immunodetection/id3/elisa. [Online] [Cited: 11 24, 2008.] http://www.millipore.com/immunodetection/id3/elisa.

Mogelsvang, Soren, *et al*. 2003. Tomographic Evidence for Continuous Turnover of Golgi Cisternae in Pichia pastoris. *Molecular Biology of the Cell.* 2003, Vol. 14, pp. 2277–2291.

Moller, Holger J. and Poulsen, Jorgen H. 1996. Staining of Glycoproteins/Proteoglycans on SDS-Gels. [book auth.] John M. Walker. *The Protein Protocols Handbook.* s.l. : Humana Press, 1996.

Morelle, W and Michalski, J-C. 2005. Glycomics and mass spectrometry. *Current pharmaceutical design .* 11, 2005, Vol. 20, pp. 2615-2645.

NCBI.
http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=nucleotide&dopt=GenBank&RID=YE1MHYP3011&log%24=nucltop&blast_rank=1&list_uids=3395509. [Online]

Philogen®. http://www.philogen.com/index.php. [Online]

Rapoport, Tom A. 2007. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature.* 2007, Vol. 450, 29, pp. 663-669.

Rossanese, Olivia W., *et al*. 1999. Golgi Structure Correlates with Transitional Endoplasmic Reticulum Organization in Pichia pastoris and Saccharomyces cerevisiae. *The Journal of Cell Biology.* 1999, Vol. 145, pp. 69-81.

Shelikoff, M, Sinskey, A. J. and Stephanopoulos, G. 1996. A modeling framework for the study of protein glycosylation. *Biotechnology and bioengineering.* 1996, Vol. 50, 1, pp. 73-90.

Umaña, P and Bailey, J. E. 1997. A mathematical model of N-linked glycoform biosynthesis. *Biotechnology and bioengineering .* 1997, Vol. 55, 6, pp. 890-908.

Vervecken, Wouter, *et al*. 2004. In Vivo synthesis of mammalian-like, hybrid-type N-glycans in Pichia pastoris. *Applied and Environmental Microbiology.* 2004, Vol. 70, 5, pp. 2639-2646.

Weerapana, Eranthie and Imperiali, Barbara. 2006. Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology.* 2006, Vol. 16, 6, pp. 91R-101R.

Wu, Chi-San. 2004. *Handbook of size Exclusion Chromatography and Related-techniques.* Second Edition. USA : Marcel Dekker, Inc, 2004.

Yan, Qiaojuan, *et al*. 2008. Characterization of a pathogenesis-related class 10 protein (PR-10) from Astragalus mongholicus with ribonuclease activity. *Plant Physiology and Biochemistry.* 46, 2008, pp. 93-99.

Young, Sandra. MALDI-TOF Mass Spectrometry. [Online] http://www.psrc.usm.edu/mauritz/maldi.html.

7. Annex

Annex A

Table 7.1 - Data from the calibration curve for BCA Protein Assay Kit with BSA as standard protein.

Concentration (mg/ml)	Absorvance average of BSA
1,0000	1,864
0,5000	1,208
0,2500	0,632
0,1250	0,360
0,0625	0,212
0,0313	0,124
0,0156	0,064
0,0078	0,033

Annex B - ELISA Technique

It was made eight ELISA plates but it is exemplified here only one plate. The Table 7.1 and Figure 7.1 are the values of the calibration curve. After that it is calculated the concentrations of the samples, in duplicate, and it is made an average of the concentrations (Table 7.3). In the end it is made the Table 7.4 and Figure 7.2 with all results.

Table 7.2 – Data from the calibration curve for this ELISA plate. Only the data in blue is used.

Dilution factor	Standard protein (mg/L)	Sample 1 Absorbance of the standard protein	Sample 2 Absorbance of the standard protein	Sample 3 Absorbance of the standard protein	Average of absorbance of the standard protein
2000	1,540	1,637	1,644	1,665	1,649
4000	0,770	1,263	1,295	1,316	1,291
8000	0,385	1,008	1,001	1,019	1,009
16000	0,193	0,878	0,903	0,870	0,884
32000	0,096	0,808	0,796	0,808	0,804
64000	0,048	0,754	0,751	0,802	0,769
128000	0,024	0,734	0,695	0,722	0,717
256000	0,012	0,734	0,702	0,696	0,711

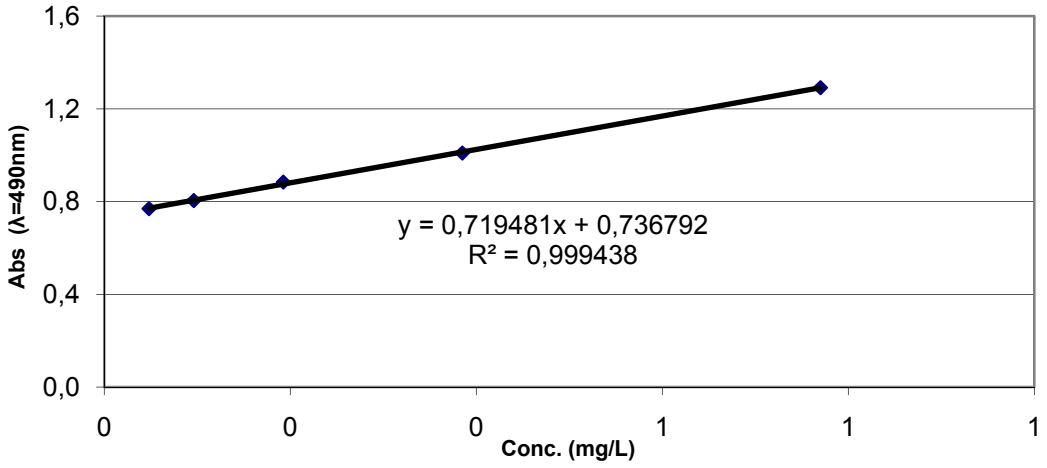

Figure 7.1 – Calibration curve for this ELISA plate.

Table 7.3– The concentrations of the product in the respective samples. The concentrations are determinated only with the values between the parameters of the calibration curve - data in blue.

	Sample 10				Sample 16			
Dilution factor	**Abs 10A**	**Abs 10B**	**Average of Abs**	**Product (mg/L)**	**Abs 16A**	**Abs 16B**	**Average of Abs**	**Product (mg/L)**
5	1,361	1,374	1,368	4,383	0,947	0,964	0,956	3,040
10	1,086	1,094	1,090	4,909	0,830	0,840	0,835	2,730
20	0,941	0,972	0,957	6,107	0,797	0,795	0,796	3,292
40	0,838	0,843	0,841	5,766	0,768	0,782	0,775	4,248
Average Concentration (mg/L)				5,594 ± 0,617	Average Concentration (mg/L)			3,020 ± 0,281

	Sample 11				Sample 18			
Dilution factor	**Abs 11A**	**Abs 11B**	**Average of Abs**	**Product (mg/L)**	**Abs 18A**	**Abs 18B**	**Average of Abs**	**Product (mg/L)**
5	1,338	1,280	1,309	3,977	0,927	0,898	0,913	2,442
10	1,037	1,007	1,022	3,964	0,790	0,791	0,791	1,493
20	0,900	0,869	0,885	4,106	0,750	0,768	0,759	1,235
40	0,814	0,782	0,798	3,403	0,734	0,736	0,735	-0,199
Average Concentration (mg/L)				3,824 ± 0,372	Average Concentration (mg/L)			1,968 ± 0,671

	Sample 13				Sample 14			
Dilution factor	**Abs 13A**	**Abs 13B**	**Average of Abs**	**Product (mg/L)**	**Abs 14A**	**Abs 14B**	**Average of Abs**	**Product (mg/L)**
5	1,283	1,304	1,294	3,869	1,219	1,257	1,238	3,483
10	0,997	0,964	0,981	3,387	1,016	1,013	1,015	3,860
20	0,856	0,861	0,859	3,383	0,853	0,832	0,843	2,938
40	0,776	0,800	0,788	2,847	0,772	0,781	0,777	2,208
Average Concentration (mg/L)				3,206 ± 0,311	Average Concentration (mg/L)			3,427 ± 0,463

	Sample 12				Sample 20			
Dilution factor	**Abs 12A**	**Abs 12B**	**Average of Abs**	**Product (mg/L)**	**Abs 20A**	**Abs 20B**	**Average of Abs**	**Product (mg/L)**
5	1,250	1,261	1,256	3,605	0,918	0,905	0,912	2,428
10	0,996	0,966	0,981	3,394	0,828	0,805	0,817	2,216
20	0,841	0,831	0,836	2,758	0,748	0,744	0,746	0,512
40	0,807	0,775	0,791	3,014	0,740	0,751	0,746	0,968
Average Concentration (mg/L)				3,252 ± 0,441	Average Concentration (mg/L)			2,322 ± 0,150

Table 7.4– All results of the concentrations of the product from the eight ELISA plates.

Sample	Age (h)	Product (mg/L)
1	4,00	
2	21,00	
3	24,00	
4	28,42	
5	46,00	**1,11**
6	48,00	**1,09**
7	51,00	**2,16**
8	54,00	**2,89**
9	57,00	**3,99**
10	69,50	**5,06**
11	72,33	**3,83**
12	75,00	**3,25**
13	78,00	**3,21**
14	81,00	**3,43**
15	93,50	**3,64**
16	96,00	**3,02**
17	99,00	**3,32**
18	102,00	**3,01**
19	105,00	**3,12**
20	117,00	**3,73**
21	120,00	**4,17**
22	123,00	**3,98**
23	126,00	**4,65**
24	129,00	**5,22**
25	141,00	**7,33**
26	143,00	**7,58**
27	145,00	**7,55**
28	147,00	**8,01**
29	149,00	**8,39**

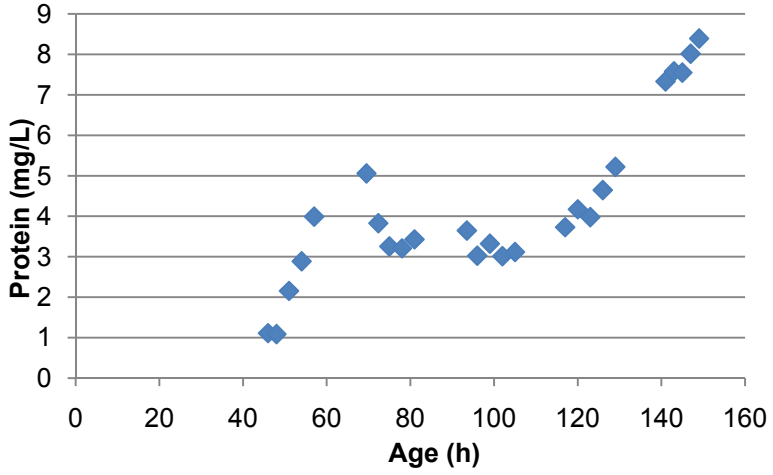

Figure 7.2 – The concentrations of the product over time.

Annex C

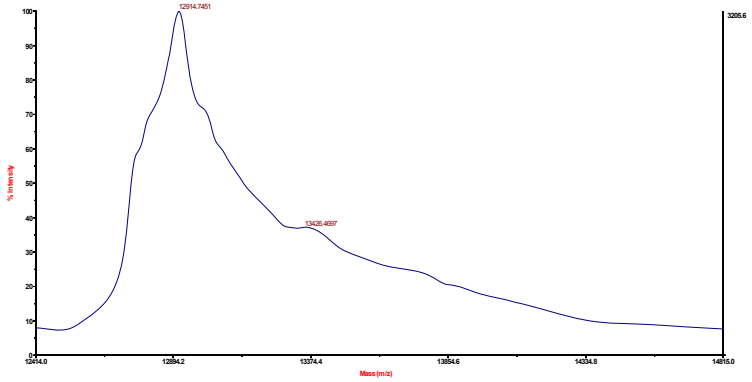

Figure 7.3 – Zoom in of the first peak of MALDI spectrum of intact protein.

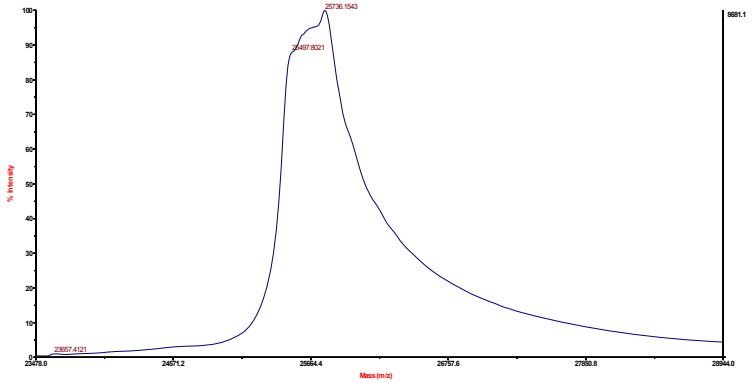

Figure 7.4 – Zoom in of the second peak of MALDI spectrum of intact protein.

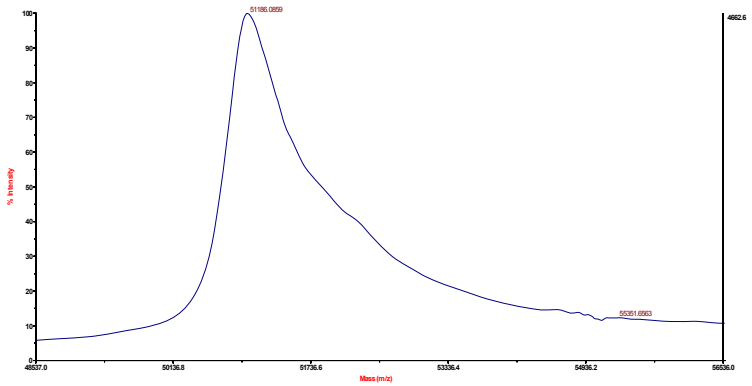

Figure 7.5 – Zoom of in the third peak of MALDI spectrum of intact protein.

Annex D

Table 7.5– Peak list of tripsin digested antibody. The peaks in blue are overlaping with the peak list os trypsin and PNGase F digested antibody, and the peaks in bolt are selected for Ms/Ms analysis[6].

Small intensity (25-190)

m/z	Intensity	m/z	Intensity	m/z	Intensity	m/z	Intensity
1128,611	140	1409,714	25	1675,778	50	1689,792	74
1877,976	50	1918,838	25	1938,990	120	1953,009	30
2018,862	102	2170,933	35	2294,027	76	2351,058	25
2523,192	36	2537,186	95				
1368,690	50	1978,476	53	2079,896	58	2609,182	26
Medium intensity (191-1900)							
1085,607	**1014**	**1142,623**	**214**	**1352,695**	**706**	**1632,773**	**1069**
1895,983	**417**	**2032,861**	**391**	**2184,934**	**391**	**2466,167**	**442**
2480,178	**624**						
High intensity (1990-3800)							
1975,851	**3800**	**2127,921**	**1840**	**2423,160**	**2737**		

Not Identified	Identified by PMF	Identified by PMF & MS/MS
1975,851	**2184,934**	**1895,983**
2032,861		**2423,160**
2127,921		**1085,607**
		1632,773
		2480,178
		2466,167

[6] Data provided/obtained by the Mass Spectrometry Laboratory, Analytical Services Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa.

Table 7.6 – Peak list of tripsin and PNGase digested antibody. The peaks in blue are overlaping with the peak list os trypsin and PNGase F digested antibody, the peaks in bolt are selected for Ms/Ms analysis and the peak in orange the potencial tryptic fragment of PNGase[7].

Small intensity (25-190)

m/z	Intensity	m/z	Intensity	m/z	Intensity	m/z	Intensity
1142,654	352/214	1409,753	93/25	1689,839	176/74	1878,038	272/50
1918,876	42/25	1953,066	305/30	2170,995	53/35	2185,006	193/391
2294,094	54/76	2351,112	51/25	2537,267	404/95		
1088,610	163	1102,620	64	1185,662	151	1199,673	54
1395,741	146	1507,693	173	1550,662	43	1706,960	33
1844,816	27	1884,901	127	1921,979	42	1927,914	46
1935,057	53	1974,914	45	1977,557	126	2001,905	117
2075,939	44	2079,977	322	2089,960	53	2123,004	72
2137,024	76	2337,145	59	2422,266	21	2449,255	72

Medium intensity (191-1900)

1085,632	**453/1014**	**1128,636**	**1683/140**	**1352,731**	**661/706**	**1632,817**	**908/1069**
1675,823	**708/50**	**1896,045**	**967/417**	**1939,057**	**2407/120**	**2018,920**	**2427/102**
2032,933	**2065/391**	**2127,900**	**518/1840**	**2423,241**	**3042/2737**	**2466,250**	**3323/442**
2480,261	**1443/624**	**2523,265**	**949/36**				
2509,261	**489/0**						

High intensity (1990-3800)

1975,851	**8292/3800**						

Not Identified	Identified by PMF	Identified by PMF & MS/MS
1975,914	**2018,920**	**1896,046**
2523,265	**2032,933**	**2423,241**
1939,057	**2480,261**	
2466,250		
2509,261		

[7] Data provided/obtained by the Mass Spectrometry Laboratory, Analytical Services Unit, Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa.

Annex E

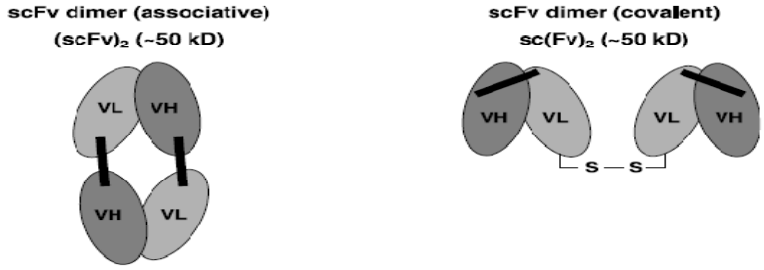

Figure 7.6 – Schematic representation of the structure of scFv dimmers. V_L – variable light domain; V_H – variable heavy domain

```
        10          20          30          40          50          60          70
EVQLLESGGG LVQPGGSLRL SCAASGFTFS SFSMSWVRQA PGKGLEWVSS ISGSSGTTYY ADSVKGRFTI

        80          90         100         110         120         130         140
SRDNSKNTLY LQMNSLRAED TAVYYCAKPF PYFDYWGQGT LVTVSSGDGS SGGSGGASTG EIVLTQSPGT

       150         160         170         180         190         200         210
LSLSPGERAT LSCRASQSVS SSFLAWYQQK PGQAPRLLIY YASSRATGIP DRFSGSGSGT DFTLTISRLE

       220         230
PEDFAVYYCQ QTGRIPPTFG QGTKVEIK
```

Figure 7.7 – Sequence of L19 (NCBI)

```
        10          20          30          40          50          60          70
EVQLLESGGG LVQPGGSLRL SCAASGFTFS SFSMSWVRQA PGKGLEWVSS ISGSSGTTYY ADSVKGRFTI

        80          90         100         110         120         130         140
SRDNSKNTLY LQMNSLRAED TAVYYCAKPF PYFDYWGQGT LVTVSSGDGS SGGSGGASTG EIVLTQSPGT

       150         160         170         180         190         200         210
LSLSPGERAT LSCRASQSVS SSFLAWYQQK PGQAPRLLIY YASSRATGIP DRFSGSGSGT DFTLTISRLE

       220         230         240
PEDFAVYYCQ QTGRIPPTFG QGTKVEIKGG GCA
```

Figure 7.8 – Aminoacid sequence of AP39. In red highlight is a possible spot for the attachment of an O-glycan.

Annex F

Table 7.7– List of glycoforms provided by GlycoMod.

Position	Sequence	Experimental Mass (m/z)	Glycoform mass (m/z)	Δmass (Dalton)	Peptide mass	Structure
68-72	FTISR	1918,8375	1296,422	-0,935	622,344	$(Man)_8$
73-76	DNSK	1921,9789	1458,475	0,289	462,207	$(Man)_9$
77-87	NTLYLQMNSLR	2001,9053	648,211	0,995	1351,692	$(Man)_4$
149-154	ATLSCR	2594,2859	1944,634	-0,676	649,322	$(Man)_{12}$
177-185	LLIYYASSR	1409,7137	324,106	0,009	1084,592	$(Man)_2$
177-186	LLIYYASSR	1895,9832	810,264	0,121	1084,592	$(Man)_5$
209-224	LEPEDFAVYYCQ QTGR	2079,9773	162,053	-0,939	1917,857	$(Man)_1$
209-225	LEPEDFAVYYCQ QTGR	2566,283	648,211	-0,791	1917,857	$(Man)_4$
225-234	IPPTFGQGTK	2018,8624	972,317	0,978	1044,56	$(Man)_6$
225-235	IPPTFGQGTK	1368,6901	324,106	-0,983	1044,56	$(Man)_2$
68-76	FTISRDNSK (1 missed cleavage)	1877,9755	810,264	0,164	1877,812	$(Man)_5$
66-72	GRFTISR (1 missed cleavage)	2294,0269	1458,475	-0,921	835,466	$(Man)_9$
66-76	GRFTISRDNSK (2 Missed cleavage)	2089,9597	810,264	-0,974	1279,663	$(Man)_5$
66-76	GRFTISRDNSK (2 missed cleavage)	1927,9136	648,211	-0,967	1279,663	$(Man)_4$
225-238	IPPTFGQGTKVEIK (1 missed cleavage)	2001,9053	486,158	0,89	1513,85	$(Man)_3$
225-243	IPPTFGQGTKVEIK GGGCA (2 missed cleavage)	2184,9336	324,106	0,859	1858,961	(Man)2

Annex G – Method for release of O-linked glycans

There are to differents ways to release the O-linked glycans.

The first one is non-reduced β-elimination: Add 1ml of NaOH (0,1M) to 10 μL of protein (1mg/ml), incubate overnight at 45ºC. Diluted the solution with 1ml of PBS at pH 7,5 and stop the reaction by add acetic acid (1M) until pH reach 4-5. Desalt the solution with a column Dowex (50W-x8(H)) 50-100mesh.

The second one is reduced β-elimination: Add 1ml of NaOH (0,1M) more $NaBH_4$ (1M) to 10 μL of protein (1mg/ml), incubate overnight at 45ºC. Diluted the solution with 1ml of PBS at pH 7,5 and stop the reaction by add acetic acid (1M) until pH reach 4-5. Evaporate the methyl-borate with 1 ml of methanol and dry at 70ºC for 5min (repeat the process three times). Desalt the solution with a column Dowex (50W-x8(H)) 50-100mesh.

Annex H – Glycosylation mathematical models code

The deterministic model is chosen when "istoch" is zero. This model solves two differential equations with ode45 function, one that concerns the Endoplasmatic Reticulum and the other the *Golgi* apparatus. The results appear: in a figure with two graphic – the first one correspond to the elimination of the free mannose and the protein concentration and the formation of protein with one mannose and the protein with two mannoses in Endoplasmatic Reticulum, and the second correspond to the elimination of the free mannose and protein with two mannoses and the formation of proteins with three, four and five mannoses; and a table with the outflow composition of the *Golgi*.

The stochastic model is chosen when "istoch" is one. This model is based on the Chemical Master Equation (CME) – a probability density function – with a numerical solution of *Monte Carlo* method with the *tau leaping* acceleration. The *Monte Carlo* method simulates the chemical evolution for an ensemble of model instances by randomly applying the reactions of the system. *Tau leaping* acceleration is pre-select time internal τ (TAU) that encompasses more than one reaction event to accelerate the system (Gillespie, 2005). This method applies the *Poisson* distribution. The results of the stochastic model appear in two distinct figures and a table. The first figure is similar to the figure in the deterministic model, the axes are the same. The second figure is composed by six histograms with a population of 100 – the first one is the distribution of the ratio between protein concentration and initial concentration of mannose; the second is the distribution of the ratio between protein with one mannose concentration and initial concentration of mannose; the third is the distribution of the ratio between protein with two mannose concentration and initial concentration of mannose; the fourth is the distribution of the ratio between protein with three mannose concentration and initial concentration of mannose; the fifth is the distribution of the ratio between protein with four mannose concentration and initial concentration of mannose; and the sixth is the distribution of the ratio between protein with five mannose concentration and initial concentration of mannose. The table has the outflow composition of the *Golgi*.

MATLAB Code:

```
function scfvglyco
%
% SCFVGLYCO
% Deterministic and stochastioc simulation of O-glycosylation
%
% function scfvglyco
%
%
% Copyright, 2009 -
% This M-file and the code in it belongs to the holder of the
% copyrights and is made public under the following constraints:
% It must not be changed or modified and code cannot be added.
% The file must be regarded as read-only.
% In case of doubt, contact the holder of the copyrights.
%
% AUTHORS:  Sara Bispo, João dias and Rui Oliveira
%
% Copyright holder:
%
% Rui Oliveira
% Biochemical Engineering Group
% Department of Chemistry
% Faculty of Sciences and Technology
% New University of Lisbon
% P-2829-516 Caparica, Portugual
% Phone  +351 212948385
% Fax    +351 212948385
% E-mail rui.oliveira@dq.fct.unl.pt
%
% $ Version 1.00 $ Date October 2009 $ Not compiled $
%
global istoch
%
% Glycosylation is modeled as a PFR reactor. The follwoing concentrations
% are at the entrance of the PFR
%
Y0(1) = 1;        % M/M0 is always 1
Y0(2) = 0.3;      % P/M0
Y0(3) = 0;        % P-M/M0
Y0(4) = 0;        % P-MM/M0
Y0(5) = 0;        % P-MMM/M0
Y0(6) = 0;        % P-MMMM/M0
Y0(7) = 0;        % P-MMMMM/M0

NMAN0=100;  %number of mannose molecules at the entrance of the PFR;
           %this parameter is only
           %important if istoch=1; it is mandatory to work with number of
           %molecules in stochistic kinetic models
           %the lower NMAN0 the bigger is the difference between stochastoc
           %and deterministic simulation

istoch=0;   % if istoch=0, deterministic simulation
            % if istoch=1, stochastic glycosylation simulation by the MONTE
            % CARLO method and TAU LEAPING acelaration. The possoin
            % distribution is used to calculate the number of times a given
            % reaction occurs in a time interval

TAU = 0.05; % time step in percentage of time residence.

%--------------------------------------------------------------------------
% DON'T CHANGE CODE FROM HERE ON
%--------------------------------------------------------------------------
TSPANRE=[0:TAU:1];  % time goes from 0 to 1 time residence
TSPANGOLGI=[0:TAU:1]; % time goes from 0 to 1 time residence
niter=istoch*99+1;
y0=Y0;
for iter=1:niter


if istoch==0
    [tret,yret] = ode45(@retend,TSPANRE,y0);
    y0Golgi=yret(end,:);
    [tGolgi,yGolgi] = ode45(@glyGolgi,TSPANGOLGI,y0Golgi);
else
    y=y0;
    yret(1,:)=y0';
    for i=1:length(TSPANRE)-1
        [yder rM A drM_dy]=retend(TSPANRE(i),y);
%         for k=1:length(rM)
%             for l=1:length(rM)
%                 fkl(k,l)=drM_dy(k,:)*A(:,l);
%             end
%             mu(k)=fkl(k,:)'*rM;
%             stdsq(k)=((fkl(k,:)').^2)*rM;
%         end
%         pause
        for j=1:length(rM)
            ntimesR(j)=poissrnd(rM(j)*NMAN0*(TSPANRE(i+1)-TSPANRE(i)));
            y(:)=y(:)+ntimesR(j)*A(:,j)/NMAN0;
        end
        y=max(y,0);
        yret(i+1,:)=y';
    end
    tret=TSPANRE';
    yGolgi(1,:)=y';
    for i=1:length(TSPANGOLGI)-1
        [yder rM A drM_dy]=glyGolgi(TSPANGOLGI(i),y);
```

```
        for j=1:length(rM)
            ntimesR(j)=poissrnd(rM(j)*NMAN0*(TSPANGOLGI(i+1)-TSPANGOLGI(i)));
            y(:)=y(:)+ntimesR(j)*A(:,j)/NMAN0;
        end
        y=max(y,0);
        yGolgi(i+1,:)=y';
    end
    tGolgi=TSPANGOLGI';
end

yfinal(iter,:)=yGolgi(end,:);

end



figure
subplot(2,1,1)
plot(tret,yret)
xlabel('(t/trmax) in the Reticulum')
ylabel('Ci/M(0) in the reticulum')
subplot(2,1,2)
plot(tGolgi,yGolgi)
xlabel('(t/trmax) in the Golgi (h)')
ylabel('Ci/M(0) in the Golgi')
legend('Location','EastOutside','Orientation','vertical','[Free Mannose]','[Protein]','[P-M]','[P-MM]','[P-MMM]','[P-
MMMM]','[P-MMMMM]')


figure
subplot(2,3,1)
hist(yfinal(:,2))
xlabel('P/M(0)')
ylabel('count')
subplot(2,3,2)
hist(yfinal(:,3))
xlabel('P-M/M(0)')
ylabel('count')
subplot(2,3,3)
hist(yfinal(:,4))
xlabel('P-MM/M(0)')
ylabel('count')
subplot(2,3,4)
hist(yfinal(:,5))
xlabel('P-MMM/M(0)')
ylabel('count')
subplot(2,3,5)
hist(yfinal(:,6))
xlabel('P-MMMM/M(0)')
ylabel('count')
subplot(2,3,6)
hist(yfinal(:,7))
xlabel('P-MMMMM/M(0)')
ylabel('count')


disp('COMPOSITION OUT OF THE GOLGI -')
disp('                  mean        std          percentage of P(0)')
fprintf(1,'M/M(0):       %12.6f %12.6f %12.6f\n',mean(yfinal(:,1)),std(yfinal(:,1)),mean(yfinal(:,1))*100);
fprintf(1,'P/M(0):       %12.6f %12.6f %12.6f\n',mean(yfinal(:,2)),std(yfinal(:,2)),mean(yfinal(:,2))/y0(2)*100);
fprintf(1,'P-M/M(0):     %12.6f %12.6f %12.6f\n',mean(yfinal(:,3)),std(yfinal(:,3)),mean(yfinal(:,3))/y0(2)*100);
fprintf(1,'P-MM/M(0):    %12.6f %12.6f %12.6f\n',mean(yfinal(:,4)),std(yfinal(:,4)),mean(yfinal(:,4))/y0(2)*100);
fprintf(1,'P-MMM/M(0):   %12.6f %12.6f %12.6f\n',mean(yfinal(:,5)),std(yfinal(:,5)),mean(yfinal(:,5))/y0(2)*100);
fprintf(1,'P-MMMM/M(0):  %12.6f %12.6f %12.6f\n',mean(yfinal(:,6)),std(yfinal(:,6)),mean(yfinal(:,6))/y0(2)*100);
fprintf(1,'P-MMMMM/M(0): %12.6f %12.6f %12.6f\n',mean(yfinal(:,7)),std(yfinal(:,7)),mean(yfinal(:,7))/y0(2)*100);

end


function [f rM A dadx]=retend(t,y)
global istoch
E1=1;
E2=1;
Man = y(1);
P0=y(2);
PM = y(3:7);

kl=1; %(tempo de residencia)^-1
kll=kl*1;

KMl=0.25;
KMll=0.1;

rM(1) = kl  * E1 *  P0 /(0.01 +  P0) * Man /(KMl +  Man);
rM(2) = kll * E2 *  PM(1)/(0.01 +  PM(1)) * Man /(KMll +  Man);
rM(3:5)= 0;

dadx(1:5,1:7)=0;
dadx(1,1) = kl  * E1 *  P0 /(0.01 +  P0) *KMl*KMl/((KMl +  Man)^2);
dadx(1,2) = kl  * E1 *  Man/(KMl +  Man) *0.01*0.01/((0.01 +  P0)^2);
dadx(2,1) = kll * E2 *  PM(1) /(0.01 +  PM(1)) *KMl*KMl/((KMl +  Man)^2);
dadx(2,3) = kll * E2 *  Man/(KMl +  Man) *0.01*0.01/((0.01 +  PM(1))^2);


% if istoch==1
%     for i=1:5
%         rM(i)=(randn(1,1)+1)*rM(i);
%     end
%     rM=max(0,rM);
```

```
% end


A =[-1 -1 -1 -1 -1     % conc manose (M)
    -1  0  0  0  0     % conc P (M)
     1 -1  0  0  0     % conc P-M (M)
     0  1 -1  0  0     % conc P-MM (M)
     0  0  1 -1  0     % conc P-MMM (M)
     0  0  0  1 -1     % conc P-MMMM (M)
     0  0  0  0  1];   % conc P-MMMMM (M)

f = A * rM';

end


function [f rM A dadx]=glyGolgi(t,y)
global istoch
E2Golgi=1;
E3Golgi=1;
y=max(0,y);
Man = y(1);
P0=y(2);
PM = y(3:7);


kll=1;
klll= kll*1;

KMll=0.1;
KMlll=0.1;

rM(1) = 0;
rM(2) = 0;
rM(3) = kll * E2Golgi *  PM(2)/(0.01 +  PM(2)) * Man /(KMll +  Man);
rM(4) = klll *E3Golgi *  PM(3)/(0.01 +  PM(3)) * Man /(KMlll + Man);
rM(5) = klll *E3Golgi *  PM(4)/(0.01 +  PM(4)) * Man /(KMlll + Man);

dadx(1:5,1:7)=0;
dadx(3,1) = kll * E2Golgi *  PM(2) /(0.01 +  PM(2)) *KMll*KMll/((KMll +  Man)^2);
dadx(3,4) = kll * E2Golgi *  Man/(KMll +  Man) *0.01*0.01/((0.01 +  PM(2))^2);
dadx(4,1) = klll *E3Golgi *  PM(3) /(0.01 +  PM(3)) *KMlll*KMlll/((KMlll +  Man)^2);
dadx(4,5) = klll *E3Golgi *  Man/(KMlll +  Man) *0.01*0.01/((0.01 +  PM(3))^2);
dadx(5,1) = klll *E3Golgi *  PM(4) /(0.01 +  PM(4)) *KMlll*KMlll/((KMlll +  Man)^2);
dadx(5,6) = klll *E3Golgi *  Man/(KMlll +  Man) *0.01*0.01/((0.01 +  PM(4))^2);

% if istoch==1
%     for i=1:5
%         rM(i)=(randn(1,1)+1)*rM(i);
%     end
%     rM=max(0,rM);
% end

A =[-1 -1 -1 -1 -1     % conc manose (M)
    -1  0  0  0  0     % conc P (M)
     1 -1  0  0  0     % conc P-M (M)
     0  1 -1  0  0     % conc P-MM (M)
     0  0  1 -1  0     % conc P-MMM (M)
     0  0  0  1 -1     % conc P-MMMM (M)
     0  0  0  0  1];   % conc P-MMMMM (M)

f = A * rM';
end
```